ENZYMES IN COG2159 OF THE AMIDOHYDROLASE SUPERFAMILY: STRUCTURE AND MECHANISM OF 5-CARBOXYVANILLATE DECARBOXYLASE (LIGW)

A Dissertation

by

ANNA V. VLADIMIROVA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Frank Raushel
Committee Members,	David P. Barondeau
	Wenshe Liu
	Margaret E. Glasner
Head of Department,	Francois Gabbai

May 2015

Major Subject: Chemistry

ABSTRACT

COG2159 of the Amidohydrolase Superfamily (AHS) is composed of a wide range of enzymes, which catalyze hydration, hydrolysis, and decarboxylation reactions. 5-Carboxyvanillate decarboxylase (LigW) belongs to COG2159 and catalyzes the conversion of 5-carboxyvanillate (5-CV) to vanillate (VAN) in the pathway for the degradation of lignin. The recombinant genes from *Sphingominas paucimobilis* SYK-6 (LigW) and *Novosphingobium aromaticivorans* DSM 12444 (LigW2) were expressed in *E. coli* in the presence of Mn^{2+} and the purified enzymes contained 1 equivalent of Mn^{2+}. The kinetic constants for the decarboxylation of 5-CV are as follows: k_{cat} = 2.0 s^{-1} and k_{cat}/K_m = 4.4 x 10^4 M^{-1} s^{-1} for the SYK-6 LigW and k_{cat} = 27 ± 1.0 s^{-1} and k_{cat}/K_m = 1.1 x 10^4 M^{-1} s^{-1} for the DSM 12444 LigW2. The pH-rate profiles are bell-shaped, these results are consistent with the required deprotonation of the invariant Asp-296 and protonation of His-226 for catalysis and/or substrate binding to occur. Alterations of LigW metal ligands significantly decrease the catalytic activity with k_{cat}/K_m values at least three orders of magnitude lower than that of the wild-type enzyme. Site-directed mutagenesis of substrate binding resides substantially lower or abolish the activity in LigW. The enzyme is also inhibited by the product vanillate, 3-methoxy-5-carboxybenzoate, and 4-hydroxy-3-methoxy-5-nitrobenzoate. The latter proved to be a tight binding inhibitor of LigW and LigW2 with a K_i^{app} = 17 nM for each of the enzymes. LigW catalyzes the exchange of the hydrogen at C-5 of VAN with deuterium over time and the product isotope effect (PIE) in a 50:50 mixture of D_2O:H_2O at pD 9.0

is 4.3. The crystal structures of LigW from SYK-6 and DSM 12444 with Mn^{2+} in the active site were determined to a resolution of 1.8 Å (PDB id: 4ICM) and 1.5 Å (PDB id: 4INF). The structure of LigW was also determined complexed with the inhibitors 5-NV, MCB, and VAN. A chemical mechanism for the decarboxylation of 5-CV to VAN by LigW has been proposed. LigW requires Mn^{2+} for catalysis, proton transfer to C-5 is likely rate limiting for the overall reaction and precedes the decarboxylation step. The protonation of the *si*-face of C5 is performed from either the hydroxyl group at C4 or the carboxylate group of Asp-296.

In addition, COG2159 γ-resorcylate decarboxylase (γ-RSD) from *Polaromonas* sp. JS666 catalyzes the conversion of γ-resorcylate to resorcinol this enzyme also catalyzes the decarboxylation of 2, 4, 6- trihydroxybenzoate, 2, 3- dihydroxybenzoate, and 2, 6-dihydroxy-4-methylbenzoate.

DEDICATION

My Darling Emiliya,

I would like to dedicate this dissertation to you! You were born at the beginning of my fourth year of graduate school. I had high blood pressure, so you arrived early. I knew with the moment I saw you that my love for you was not comparable to anything I have ever felt before. I want you to know that your dad kept it together above most, while you were in the NICU and without him and his support for the past 5 years, I won't be writing this today. So make sure you have a loving and supportive person next to you one day. I want you to know that anything is possible in life and you can always reach your goals by applying yourself and being persistent. Your grandmother brought me to the United States, when I was 16 years old. I have come a long way, since that time. Hard work always pays off, but make sure you always have a goal in mind. Most importantly, be positive and happy. I hope you will have a passion for science like me and your father, but no matter what you decide to do in life, I will always be proud of you. I love you very much!

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Frank M. Raushel, for his patience, guidance, and support throughout the course of this research. I would like to thank my committee members, Dr. David P. Barondeau, Dr. Wenshe Liu, and Dr. Margaret E. Glasner.

I would like to thank my family and friends for believing in me! I would like to thank, Dr. Merlin Eric Hobbs, for his patience and love. I would like to thank my mother, Dr. Margarita Angelova Tutsova-Trauth, for giving me the inspiration to be a scientist and encouraging me to always strive for excellence. I would like to extend my appreciation to the present and past members of the Raushel Lab for providing me with a collaborative and supportive working environment.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Advances in genome sequencing have provided a large number of protein sequences in archival databases such as UniProtKB and GenBank. In October 2014, UniProtKB/TrEMBL, an automatically curated database showed ~ 89 million protein sequences of which ~550,000 have been manually annotated and deposited into UniProtKB/Swiss-Prot (*1*). Concomitant with the rising number of protein sequences, propagation of annotation error has become prominent throughout these databases. By heavily relying on high-throughput computational predictions without experimental support, a large number of protein homologues are incorrectly annotated (*2*). Therefore, identifying a reliable molecular function, validated by biochemical experimentation of uncharacterized proteins discovered in bacterial genomes, is important. It is well known that prokaryotic genes are typically clustered together, when involved in a specific biosynthetic pathway (*3*). Analysis of the genome/operon context can thus become a convenient tool in predicting the function of a given gene product. However, in some cases, metabolic genes are not necessarily organized in an operon; therefore, other biochemical tools are used to alleviate the problem of misannotations (*4*). Convergence of sequence, structural, and mechanistic similarities of functionally characterized proteins could be instrumental in determining the function of unknown homologous (*5*). It has been estimated that 40% of seven well-characterized families of enzymes within the amidohydrolase superfamily (AHS) were incorrectly annotated in automatically curated databases (*2, 6*). Thus, by employing a combinatorial strategy, such as

genome/operon context, sequence similarity networks, analysis of known X-ray crystal structures, and utilizing various mechanistic probes, enzymatic mechanisms and functions in the AHS could be deciphered.

The AHS, identified in 1997 by Holm and Sander, is predominately composed of hydrolytic enzymes catalyzing the cleavage of C-O, C-N, and P-O bonds at the carbon or phosphorus centers of amino acids, nucleic acids, sugars, and organophosphate esters. Other divergent superfamily members include isomerases, decarboxylases, and hydratases (*7*, *8*). Prominent structural characteristics of this superfamily include mononuclear or binuclear metal centers coordinated by conserved residues at the C-terminus of a distorted $(\beta/\alpha)_8$ -TIM barrel structural fold (**Figure 1.1**). Six residues are thought to be important for catalysis and binding of 1 or 2 metals. These include an invariant aspartate from β-strand 8, which activates the hydrolytic water for catalysis. Metal binding residues for a mononuclear metal center typically include an HxH motif from β-strand 1 and two histidine residues from β-strands 5 and 6, in addition to a water molecule as seen in the AHS members adenosine deaminase (ADA) and cytosine deaminase (CDA) in **Figure 1.1A** (*7*). One of the histidine residues in the HxH motif sometimes interacts with the metal bound water and a more uncommon glutamate residue from β-strands 3 is directly ligated to the divalent metal center as observed in acetyl glucosamine-6-phosphate deacetylase (NagA) in **Figure 1.1B** (*9*). These residues are also associated with binding of two metal centers, in addition to a bridging hydroxide and a glutamate from β-strand 4 represented in phosphotriesterase (PTE), dihydroorotase (DHO), and urease, (URE, **Figure 1.1C)** (*10, 11, 12*). Furthermore, the superfamily can be divided into three sub-types based on the number of metal centers and active site

architecture (*12*, *13*). The first sub-type is characterized by enzymes that house 2 metals in the active site, one more buried site (α-site) and another more solvent exposed site (β-site). The second and third sub-types are members with a mononuclear center in either the α or β positions, respectively. Evidence of enzymes with no metal center or 3 metal centers in the AHS have been recently reported (*14*, *15*). Besides conservation of active site architecture, little sequence identity is observed among the AHS members. The sequence variation is seen in the loop regions of the $(\beta/\alpha)_8$ -TIM barrel, which are thought to undertake conformational changes essential for substrate binding and orientation (*5*, *7*). These loops vary in length and amino acid composition and serve as stereoselective discriminators and gatekeeprs to correctly align the substrate into the active site.

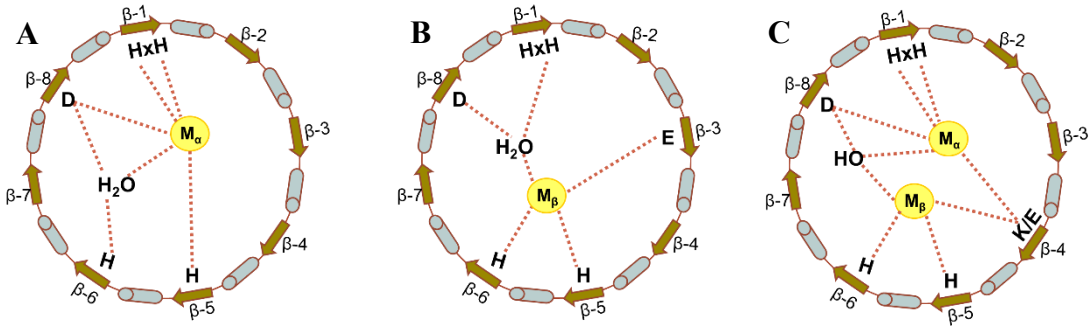
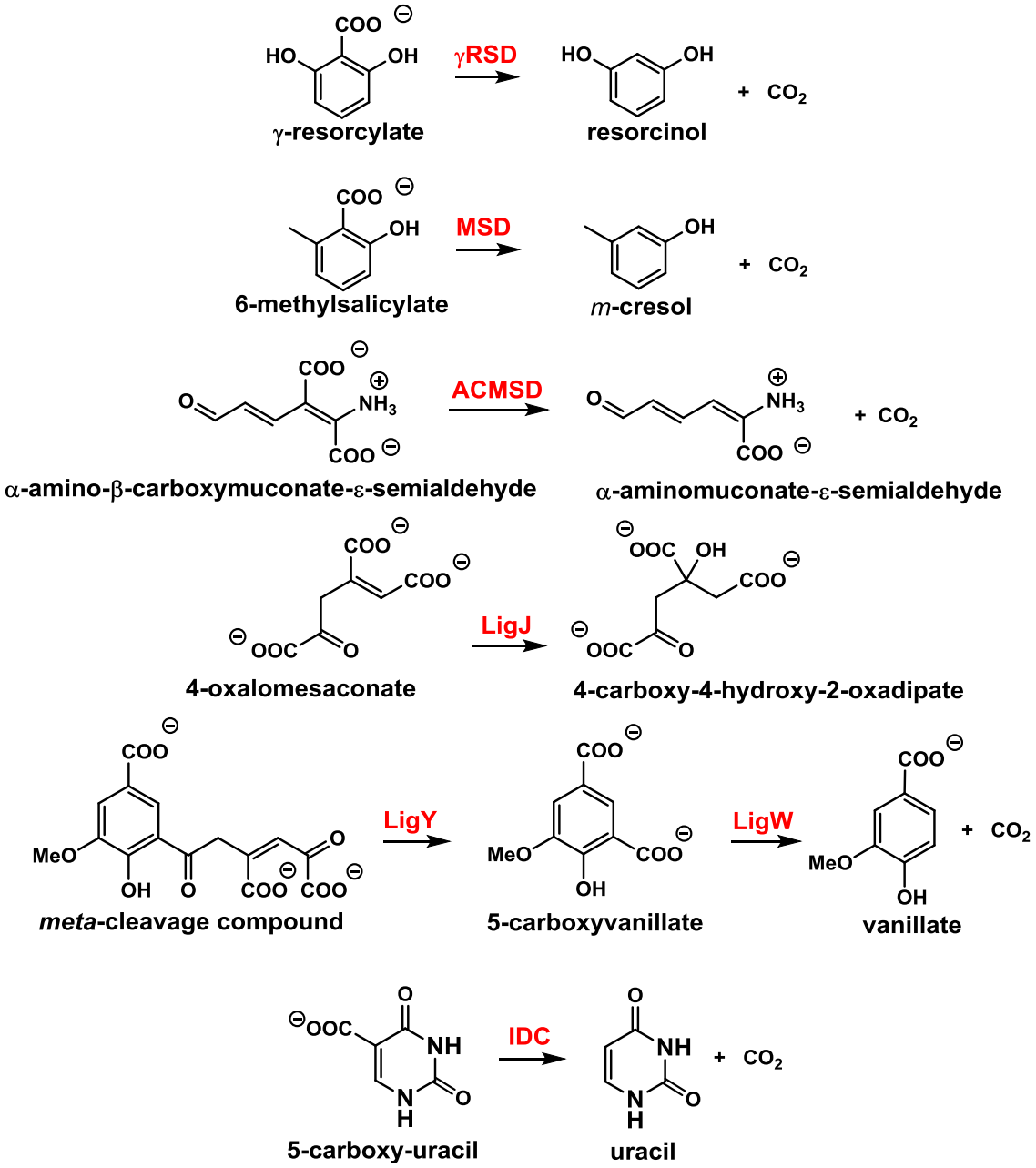


Figure 1.1: Variations of metal binding properties and active site construction of characterized AHS enzymes. Adenosine deaminase (ADA) and cytosine deaminase (CDA) (**A**), N-acetyl glucosamine-6-phosphate deacetylase (NagA) (**B**), and phosphotriesterase (PTE), dihydroorotase (DHO), and urease, (URE) (**C**).

There are 42 different reactions characterized within the AHS, which contains 36,690 members represented by 24 clusters of orthologous groups (COGs) in the NCBI database. These COGs are phylogenetically classified proteins encoded by completely sequenced genomes that are thought to be direct evolutionary descendants from a common ancestor (*16*). Members of COG2159 are the only known AHS members that catalyze decarboxylation and hydration reactions. From the 1600 non-redundant protein sequences in COG2159, the functions of 7 gene products are known as shown in **Scheme 1.1**: γ-resorcylate decarboxylase (γRSD) (*17*), α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) (*18*), OH-DDVA *meta*-cleavage compound hydrolase (LigY) (*19*), 4-oxalomesaconate hydratase (LigJ) (*20*), and 5-carboxyvanillate decarboxylase (LigW) (*21*), uracil-5-carboxylate decarboxylase (IDC) (*22*), and 6-methylsalicylic acid decarboxylase (MSD) (*23*). The crystal structures of γ-RSD (*24*), ACMSD (*25*), IDC (*22*), and LigJ are known. Structural alignments of characterized COG2159 decarboxylases have revealed that the homologs indeed share similar structural topology (**Figure 1.2A**) and overall active site architecture with a mononuclear α-site (**Figure 1.2B**), which includes conserved residues important for metal binding.



Scheme 1.1: COG2159 reactions catalyzed by known decarboxylases, hydrolases, and hydratases.

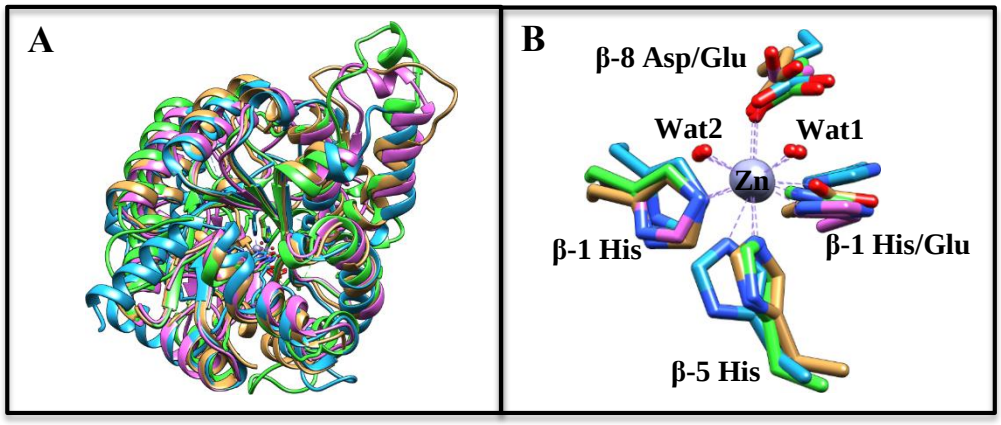


Figure 1.2: Monomeric superimposed topology of structurally characterized proteins in COG2159 of AHS. γ-Resorcylate decarboxylase (γRSD) from *Rhizobium* sp. MTP-10005 in beige (PDB id: 2DVT), α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) from *Pseudomonas fluorescens* in purple (PDB id: 2HBV), 4-oxalomesaconate hydratase (LigJ) from *Rhodopseudomonas palustris* CGA009 in blue (PDB id: 2GWG), uracil-5-carboxylate decarboxylase (IDC) from *Cordyceps militaris* in green (PDB id: 4HK5) (**A**); The respective active site superposition of the characterized COG2159 enzymes with the corresponding color code as described in (**A**). The ExH motif seen in γRSD from β-1 (beige) deviates from the typical HxH seen in ACMSD (purple), LigJ (blue), and IDC (green). Highly conserved His from β-5 and Asp from β-8 seen in all the proteins with exception of Glu replacing the Asp from β-8 in the hydratase LigJ. Together with two metal bound H_2O molecules (Wat 1 and 2), the Zn metal is in an octahedral geometry with exception of γRSD, which is missing Wat1 retaining a trigonal bipyramidal geometry.

To further examine conserved residues important for catalysis in COG2159 decarboxylases and hydratases, sequence alignments of structurally characterized proteins with enzymes of known function have revealed that the HxH motif from β-strand 1 is not strictly conserved in all the members (**Figure 1.3)**. Nevertheless, it is seen in ACMSD from *Pseudomonas fluorescens* (gi|116667628, PDB id: 2HBV), IDC from *Metarhizium anisopliae* (gi|541881569, PDB id: 4HIJ), MSD from *Aspergillus clavatus* (gi|401829650), LigJ from *Rhodopseudomonas palustris* CGA009 (gi|39937755, pdb id: 2GWG), and LigY from *Sphingomonas paucimobilis* (gi|4520316). The alternate motifs also play a crucial role in binding of the metal center in LigW and γRSD. These catalytic residues include: ExA and ExH motifs from β-strand 1 seen in LigW proteins from *Sphingomonas paucimobilis* SYK-6 (gi|19110430) and *Novosphingobium aromaticivorans* DSM 12444 (gi|87198821) and γRSD enzymes from *Rhizobium* sp. MTP-10005 (gi|116667102, pdb id: 2dvt) and *Polaromonas* sp. JS666 (gi|91787937), respectively. The metal ligands, histidine from β-strands 5 and aspartate from β-strand 8 are all conserved in all the members. However, it is observed that the only functionalized hydratase (LigJ) and hydrolase (LigY) contain a glutamate in β-strand 8 in comparison to all the known decarboxylases in COG2159. In addition, histidine from β-strand 6 is highly conserved among all functionally known COG2159 enzymes (**Figure 1.3**).

Active site residues in ACMSD, γRSD, and IDC possibly responsible for catalysis and substrate binding have been identified based on crystal structures of ACMSD in complex with 1, 3-dihydroxyacetonephosphate (DHAP) (PDB:

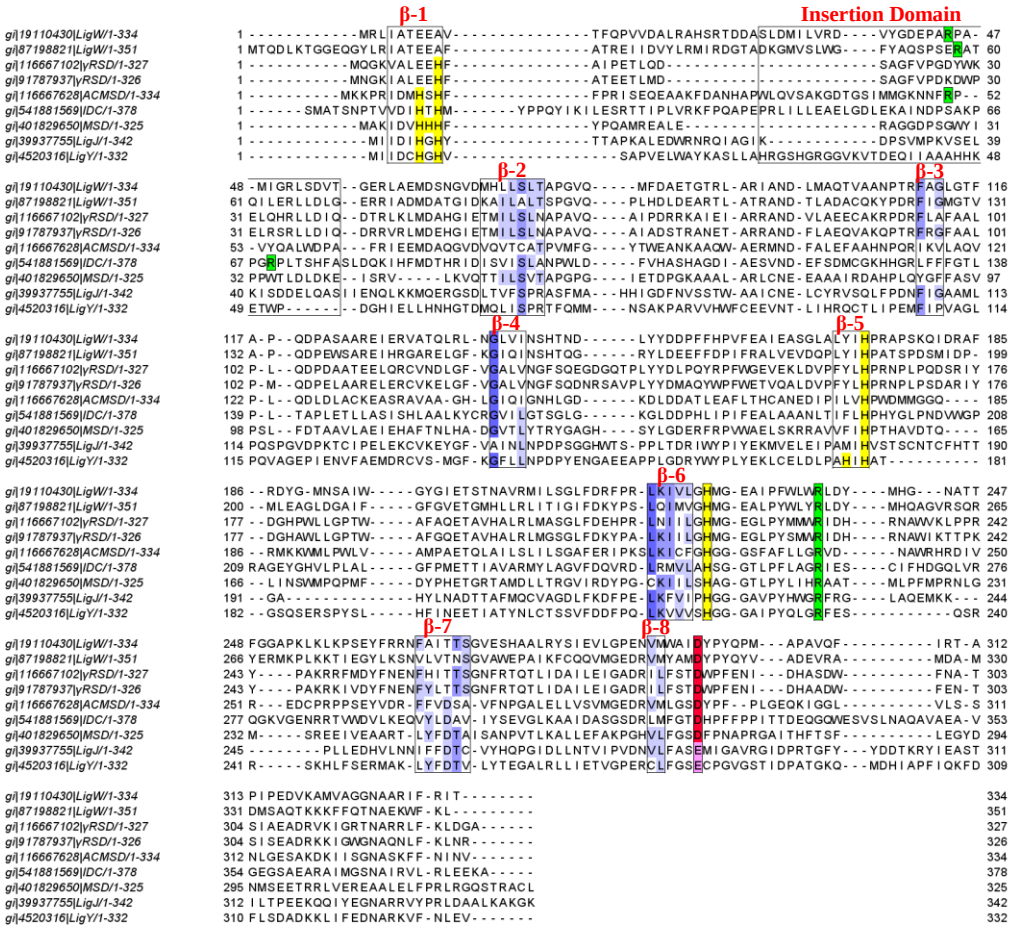


Figure 1.3: Sequence alignments of structurally and functionally characterized COG2159 decarboxylases, hydratases, and hydrolases with β-strands boxed and labeled in red. Conserved histidine residues are seen in yellow from β-strands 1, 5, and 6. Highly conserved arginine residue (green) following β-strand 6 has been shown to be important for substrate binding and orientation. The aspartate from β-strand 8 is seen in all the decarboxylases and it is replaced by a glutamate in LigJ and LigY. An insertion domain is seen in the structurally characterized decarboxylases boxed and labeled in red. Each domain contains an arginine residue (green) proposed to be important for coordinating additional substituents of the substrate molecule in ACMSD and IDC.

2WM1) shown in **Figure 1.4C** and pyridine-2, 6-dicarboxylic acid (PDC) (pdb id: 4IH3), γRSD in presence of the substrate 2, 6-dihydroxybenzote (pdb: 2DVU) represented in **Figure 1.4B** and the inhibitor 2, 6-dihydroxybenzaldehyde (pdb: 2DVX), and IDC in complex with inhibitor 5-nitrouracil (pdb: 4HK6) illustrated in **Figure 1.4A** and product uracil (pdb: 4HK7) (*22*, *24*, *25*, *27*). These include an arginine residue typically found in the insertion domain between β-strand 1 and α-helix 1 in ACMSD and IDC responsible for orientation and stabilization of the substrate into the active site (**Figure 1.5**) (*22*, *26*). An arginine in this position is not seen in γ-RSD, since the γ-resorcylate substrate lacks additional charged substituents at the *meta* or *para* positions on the phenyl ring.

Additionally, it has been established that the dimeric form is necessary for catalysis in these enzymes, since the highly conserved arginine residue from the adjacent subunit has been found to interact with the leaving carboxyl group in γ-RSD (*24*), ACMSD (*27*), and IDC (*22*) (**Figure 1.5**). The metal binding residues in γ-RSD include Glu-8 and His-10 from β-strand 1 (ExH motif) along with a His-164 from β-strand 5 and an Asp-287 from β-strand 8. This enzyme is Zn^{2+} dependent and catalyzes the decarboxylation of an *ortho*-substituted hydroxybenzoic acid (*24*). In an attempt to identify the functional roles of the catalytic residues in γ-RSD and the other decarboxylases from COG2159, the crystal structure of the substrate-bound complex of γ-RSD is further examined (**Figure 1.4B)**. The structure of γ-RSD, complexed with γ-resorcylate, reveals that the divalent cation is ligated to the carboxylate oxygen of the substrate in a trigonal bipyramidal geometry.

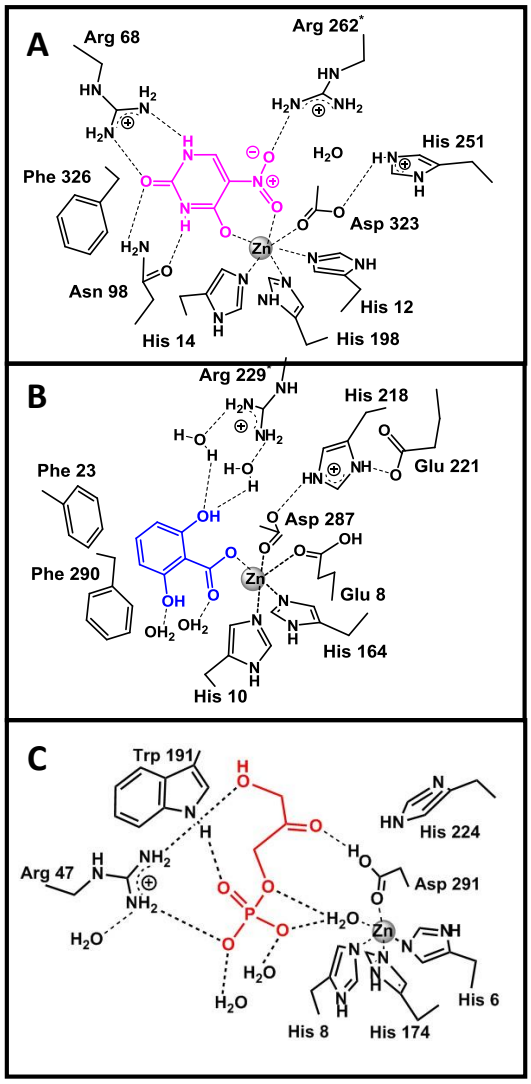


Figure 1.4: Active sites structures of representative decarboxylases from cog2159. (**A**) γ-RSD complexed with resorcylate (PDB id: 2DVU). (**B**) IDC complexed with 5-nitrouracil (PDB id: 4HK6). (**C**) Proposed model for the binding of α-aminomuconate-ε-semialdehyde in the active site of ACMSD.

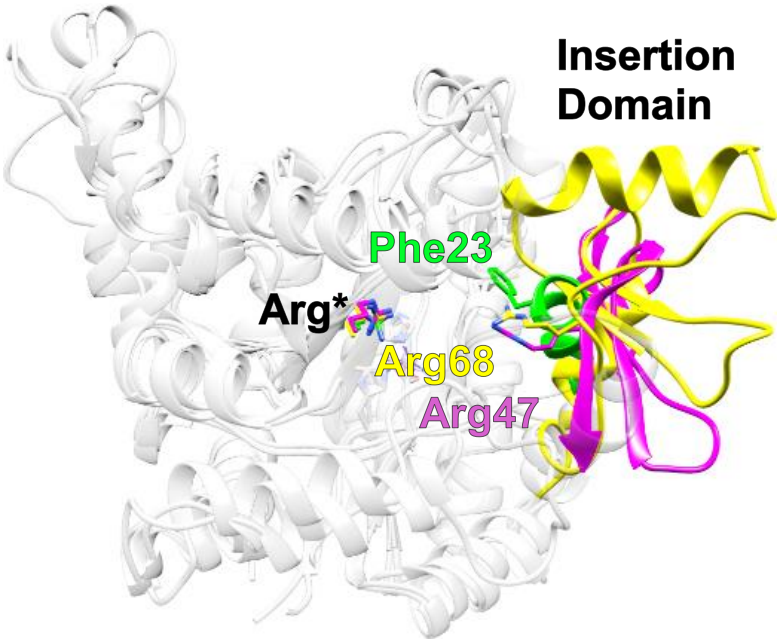
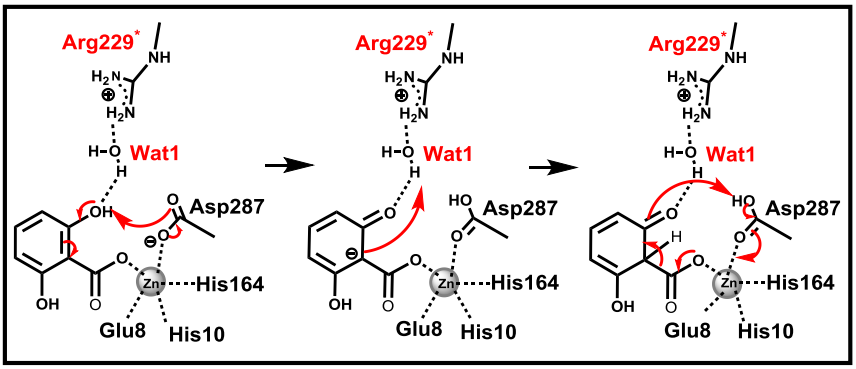


Figure 1.5: Subunit A ribbon superposition of the structurally characterized ACMSD (PDB id: 4IH3) in presence of pyridine-2,6-dicarboxylic acid, IDC (PDB id: 4HK6) complexed with 5-nitro uracil, and γRSD (PDB id: 2DVU) with substrate γ-resorcylate. Insertion domains and critical catalytic residues shown in ACMSD, two arginine residues and a β-sheet in purple; IDC, two arginine residues and the insertion domain composed of two helices and a β-hairpin in yellow; γRSD, an arginine and a phenylalanine from the inserted α-helix in green. Arg* indicates the highly conserved Arg residues from the adjacent subunits, which help stabilize the leaving carboxylate of the substrates in all the proteins. The residues on the insertions domains are Arg47 in ACMSD, Arg68 in IDC, and Phe23 in γRSD.

A hydrogen bond triad is formed between Asp-287 from β-strand 8, His-218 from β-strand 6, and a Glu-221 residue from another subunit of the γRSD tetramer. Furthermore, Phe-290 is thought to stabilize γ-resorcylate through π-π stacking interactions and together with Phe-189 limits the substrate to having a planar geometry. These hydrophobic residues are proposed to be essential for substrate orientation into the active site. A hydrophilic cluster formed between two water molecules, Asp-287 from β-strand 8, Arg-229 and Asn-234 residues from another subunit of the γRSD tetramer are thought to be important for interaction with the substrate 2-hydroxy group (**Figure 1.4B**). Based on the available crystal structures of γ-RSD complexed with substrate γ-resorcylate (pdb id: 2dvu) and inhibitor 2, 3-dihydroxybenzaldehyde (pdb: id 2vx), a chemical mechanism for the decarboxylation of γ-resorcylate has been proposed (**Scheme 1.2**). Upon access of the substrate into the active site, γ-RSD changes its conformation to accommodate the substrate γ-resorcylate. Superposition of the native



Scheme 1.2: Proposed catalytic mechanism of γ-RSD based on PDB id: 2DVU crystal structure.

enzyme (PDB id: 2DVT) with other characterized COG2159 decarboxylases (**Figure 1.5**), reveals the insertion domain between β-strand 1 and α-helix 1. The insertion domain is not resolved in all the γ-RSD crystal structures. Residues 20 to 24 of the 372 residue monomers C and D of γ-RSD homotetramer in the native enzyme structure (PDB id: 2DVT) are missing. These amino acid residues belong to an α-helix in the insertion domain and contain Phe-23 seen in domains A and B. This residue could be important for the stabilization of the substrates phenyl ring through T-shaped stacking interactions. Amino acid omissions from the inserted α-helix are also seen in the inhibitor (PDB id: 2DVX) and substrate (PDB id: 2DVU) bound crystal structures. The α-helix loop changes its main and side chain conformations, so that Ser-20 and Phe-23 encapsulate the substrate into the active site. Phe-23 sustains a van der Waals interaction with a shortest distance of 3.5 Å from the γ-resorcylate benzene ring.

Upon access of the substrate into the active site, one of the oxygens of the carboxyl group acts as a monodente ligand to the metal center and C-2 hydroxyl group is stabilized by the metal bound Asp-287. The latter is further polarized by His-218 and Glu-221, thus forming a hydrogen bond triad. Further, Asn-234 and Arg-229 from the adjacent subunit of the γ-RSD homotetramer act as a recognition site for the substrate carboxyl group and adjacent hydroxyl group. The decarboxylation reaction of γ-RSD is initiated by the invariant Asp-287, which abstracts the proton from the C-2 hyrdoxyl group followed by formation of a carbanion at the C-1 of the phenyl ring. A water molecule (Wat1) is at a distance of 3.76 Å from the C-1 and interacts with His-218 and Arg-229; thus, lowering Wat1 pK_a value. Once Asp-287 is protonated, the hydrogen bond interaction with His-218 is weakened, which in turn reduces the interaction with

Wat1. Therefore, it is proposed that the Wat1 would serve as an active site acid adding one of its protons to C-1 and forming sp^3 hybridized intermediate. Decarboxylation at C-1 occurs with concomitant abstraction of the Asp-287 proton by the phenolate, resulting in the formation of CO_2 and resorcinol (**Scheme 1.2**).

Although the proposed mechanism is plausible, several questions still remain concerning the inability of the crystalline enzyme to decarboxylate the substrate. In addition, Wat1 is presumably polarized by Arg-229 (2.6 Å) and further it is at a shorter distance (3.1 Å) from the C-2 in comparison to C-1 (3.8 Å) of the benzene ring; thus, it could nucleophilic attack at C-2 with a subsequent protonation at C-1 by another conjugate acid. Decarboxylation than can occur with elimination of H_2O. To date, γ-RSD chemical mechanism has not been fully elucidated and the information obtained by the crystal structures give limited knowledge for other COG2159 enzymatic mechanisms.

To further investigate the substrate specificity and mechanisms of enzymes in COG2159, human ACMSD crystal structure in presence of the inhibitor 1, 3-dihydroxyacetonephosphate (DHAP) (PDB id: 2WM1) and pyridine-2, 6-dicarboxylic acid (PDC) (PDB id: 4IH3) are analyzed in **Figure 1.6 A and B**, respectively (*24*, *27*). Superposition of the two structures showed that ACMSD is a monomer with 1, 3-dihydroxyacetonephosphate (DHAP) in the active site and it possesses all conserved metal ligands necessary to bind a mononuclear metal center as described previously. Further analysis of the ACMSD active site reveals a bound Zn^{2+} ion forming a distorted trigonal bipyramidal geometry with His-6, His-8, His-174, Asp-291, and a water molecule (**Figure 1.6**). The ACMSD inhibitor DHAP binding site consists of an Arg-

47, which establishes a salt bridge with the phosphate moiety and hydrogen bonds to the keto group. Additionally, Arg-235 from the α6 helix in ACMSD is conserved as determined by sequence alignments with COG2159 proteins and it is seen in the PDC bound structure.

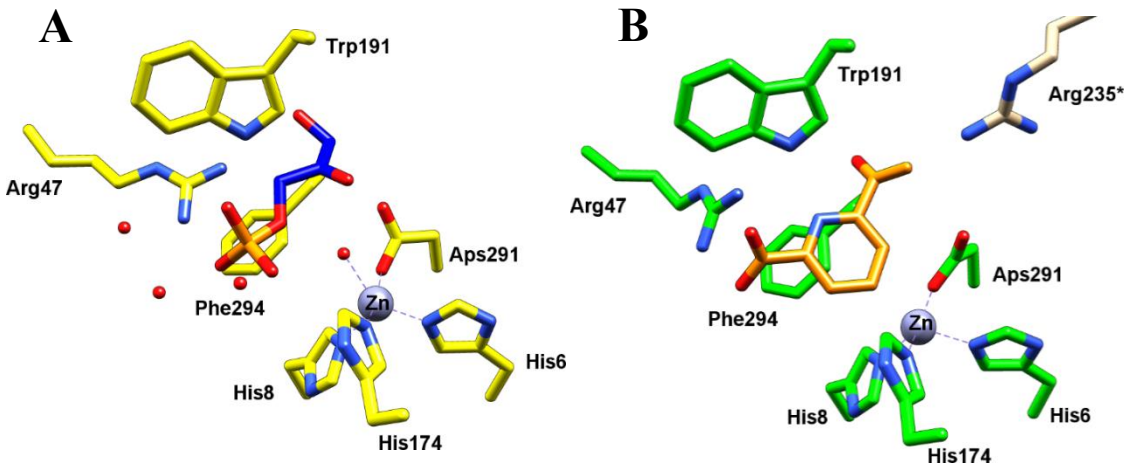
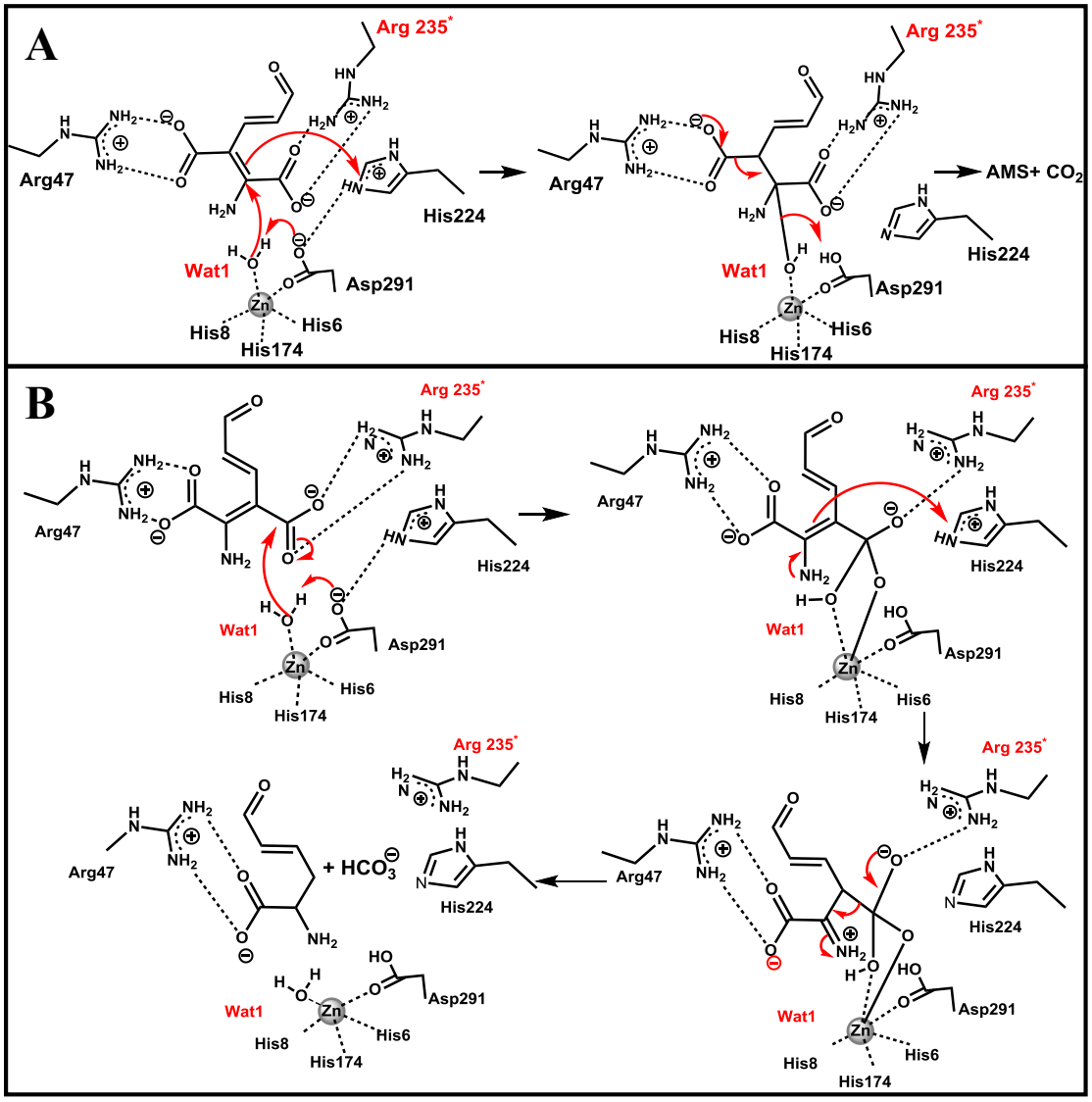


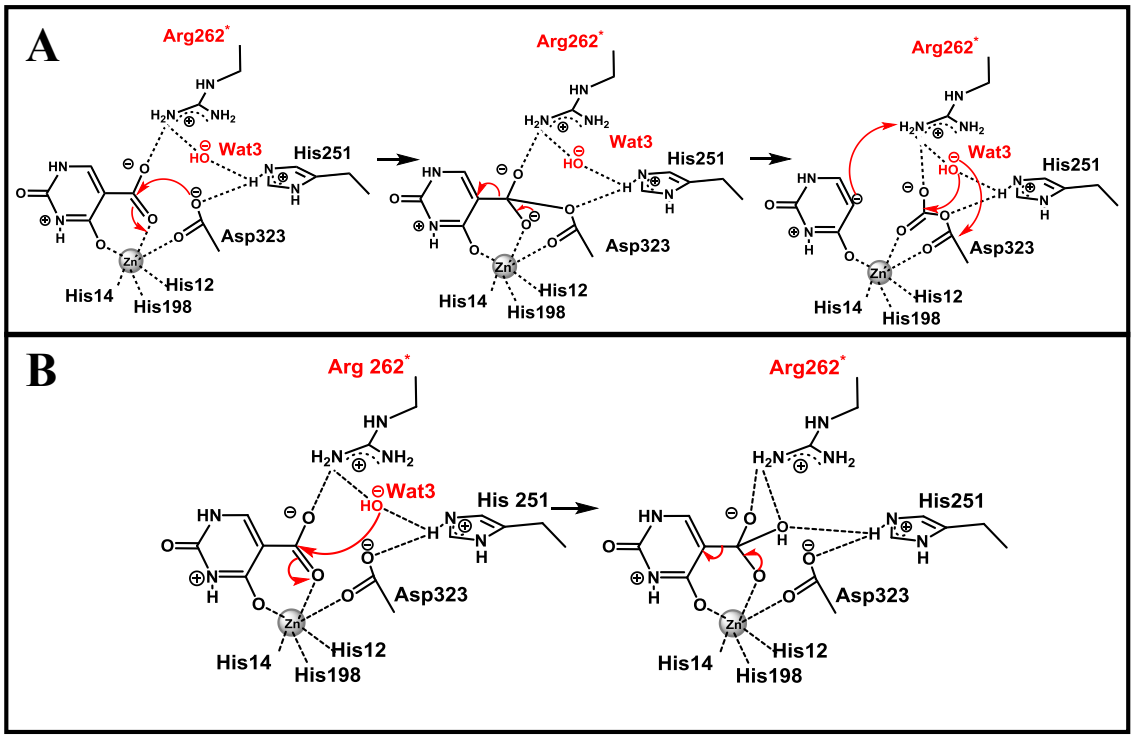
Figure 1.6: Active site of human ACMSD with two inhibitors bound 1, 3-dihydroxyacetone-phophate (DHAP) in blue (**A**) and pyridine-2, 6-dicarboxylic acid (PDC) in orange (**B**). Arg235* in khaki extends from the adjacent monomer of the PDC bound structure.

Unfortunately, the structure of this enzyme has not been determined in the presence of an inhibitor or other ligand that more resembles the substrate or the decarboxylated product. However, several reaction mechanisms have been proposed as seen in **Scheme 1.3**. In the first mechanism, ACMS presumably binds into the active site with the C-2 carboxyl group oriented towards the Zn^{2+} ion with the leaving carboxylate group stabilized by Arg51, a water molecule (Wat1) is than activated by Asp-291. A nucleophilic attack by Wat1 is then achieved at C-2 with subsequent protonation at C-3 by the protonated form of His-224 (*71*). In addition, recent ACMSD mechanism models suggest that the histidine from β-strand 6 would activate a metal bound water, which will add to C-2 with protonation at C-3 (*71*). Decarboxylation of the C-3 carboxyl group is initiated with collapse of the tetrahedral intermediate and release of CO_2 and 2-aminomuconate semialdehyde (AMS) (Figure 1.3A). In the second mechanism, ACMS binds into the active site with the carboxyl group at C-3 oriented towards the metal center and forming ionic interactions with Arg235* from the adjacent subunit. Thus, Arg51 from the insertion domain would interact with the C-2 carboxyl group instead. The metal bound Asp291 abstracts the proton from Wat1 resulting in a nucleophilic attack by the hydroxide on the carbon center of the C-3 carboxyl group forming unstable tetrahedral intermediate. The lone pair of the amino group is delocalized and protonation at C3 is achieved by His224. The HCO_3^- group must be coordinated by the Zn^{2+} ion for the decarboxylation step to occur. Exchange of the metal ligated bicarbonate with a water molecule, results in the dissociation of AMS from the active site (**Figure 1.3B**).



Scheme 1.3: Proposed catalytic mechanisms of ACMSD. (**A**) Nucleophilic attack by the activated metal bound H_2O at the C-2 with subsequent protonation at C-3 of the substrate (ACMS) followed by decarboxylation. (**B**) Nucleophilic attack of Wat1 at carbon of the C-3 leaving carboxylate group followed by delocalization of the electron pair of the amino group and protonation at C-3. Collapse of the tetrahedral intermediated results in decarboxylation at C-3.

When the mechanisms of ACMSD and IDC are compared, it is seen that the zinc-dependent isoorotate decarboxylase (IDC) presumably utilizes a non-metal bound water for the decarboxylation reaction. The high-resolution crystal structure of IDC was determined previously in the presence of 5-nitro-uracil as presented in **Figure 1.4A** (PDB id: 4HK6). In IDC, the active site zinc is coordinated to the nitro group and the adjacent oxygen of the bound inhibitor. In addition, Arg-262* from the adjacent subunit interacts with the nitro-substituent of the inhibitor in a similar manner to that of Arg-235* in ACMSD. In the first proposed mechanism, the substrate presumably binds into the active site replacing the two metal bound waters with the C-4 hydroxyl oxygen and C-5 carboxylate oxygen in a bidentate fashion. In addition, His-12, His-14, H-195, and Asp-323 are ligated to the Zn^{2+} ion establishing an octahedral geometry. The O-2 and the N-3 form hydrogen bonds with Asn-98. Whereas, N-1 is coordinated by Arg-68, which extends from the insertion domain (**Figure 1.4**). One of the oxygen atoms of Asp-323 is activated by the metal center, while the second oxygen is hydrogen bonded to His-251. A nucleophilic attack by Asp-323 is performed on the carbon center of the C-5 carboxyl group. The unstable tetrahedral intermediate collapses concomitantly with protonation at C-5 from solution forming uracil and a carboxylated aspartate intermediate (**Figure 1.5A**). The second mechanism, involves the addition of the hydroxide to the carbon center of the leaving carboxylate, forming an unstable tetrahedral intermediate, which is stabilized by Arg-262* and the metal center. Collapse of this intermediate leads to the formation of the product uracil and HCO_3^- (**Figure 1.5B**).



Scheme 1.4: Proposed catalytic mechanism of IDC based on the crystal structure with inhibitor 5-nitor uracil (5-NV, PDB id: 4HK6). Nucleophilic attack of hydroxide on two possible carbon centers of the carboxyl aspartate intermediate. (**A**) Involving the formation of covalently bound tetrahedral intermediate with Asp323. (**B**) Addition of H_2O to the carbon center of the C-5 carboxyl group of the substrate, followed by decarboxylation.

From the 7 functions in COG2159, the mechanisms of the gene products LigJ, LigW, and LigY in *Sphingomonas paucimobilis* SYK-6 together with MSD from *Aspergillus clavatus* have not been elucidated. Preliminary information on γ-RSD, ACMSD, and IDC could be used to decipher the catalytic mechanism of known COG2159 members and further aid in the characterization of unknown homologous. To help visualize the phylogenetic relationship between genes in COG2159, sequence similarity networks (SSN) have been generated using the interactive software Cytoscape. This tool allows for a comprehensive view of related genes based on pairwise sequence identity at a given stringent threshold or BLAST E-value (*28*). At high stringency, sequences represented as nodes are separated into discrete clusters. Highly coalesced clusters share higher sequence similarity in comparison to nodes with fewer and longer connections (*29*).

For COG2159, an E-value cutoff of 10^{-70} was arbitrarily chosen to differentiate between clusters by assigning group numbers, which can be further traced when the stringency is relaxed and groups become interconnected as seen in an E-value of 10^{-20} (**Figure 1.7**). The purpose of Cytoscape mapping is to visualize the relationship between orthologs and orthologous sets of paralogs of known and unknown functions in the COG2159 network at various stringencies (*29, 30*). Since orthologs are genes from different species that have evolved from a common ancestor, typically the same function is retained throughout evolution (*31*). In understanding the relationship between orthologs, one could predict new functions in various genomes and gain better understanding of the gene organization within an organism by exploring the genomic context (*32*). Thus, the manipulation of COG2159 through sequence similarity networks

(SSN), could be central to the meaningful comparison of enzymatic functions. Utilization of this methodology begins with analysis of the sequence similarity map at E-value of 10^{-70} (**Figure 1.7**), which reveals that group 1 includes all γRSD and LigW proteins, whereas groups 2, 17, 26, 30, and 31 represent ACMSD, LigJ, MSD, LigY, and IDC, respectively. When the stringency is relaxed to 10^{-20}, it is observed that all the known functions in COG2159 are gathered in a large cluster; whereas, the rest of the clusters remain occupied by unknown functions.

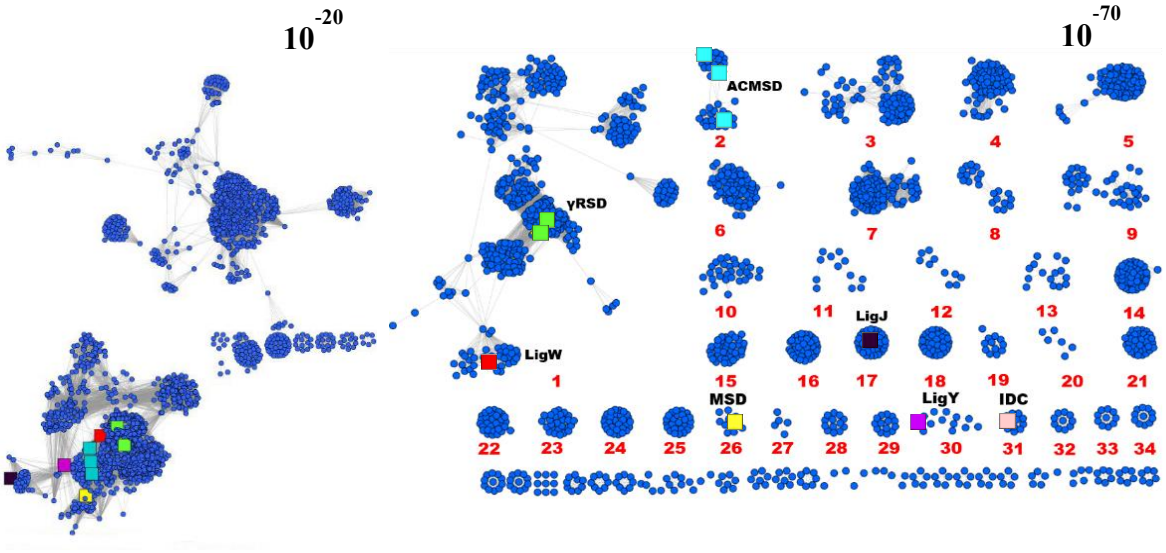
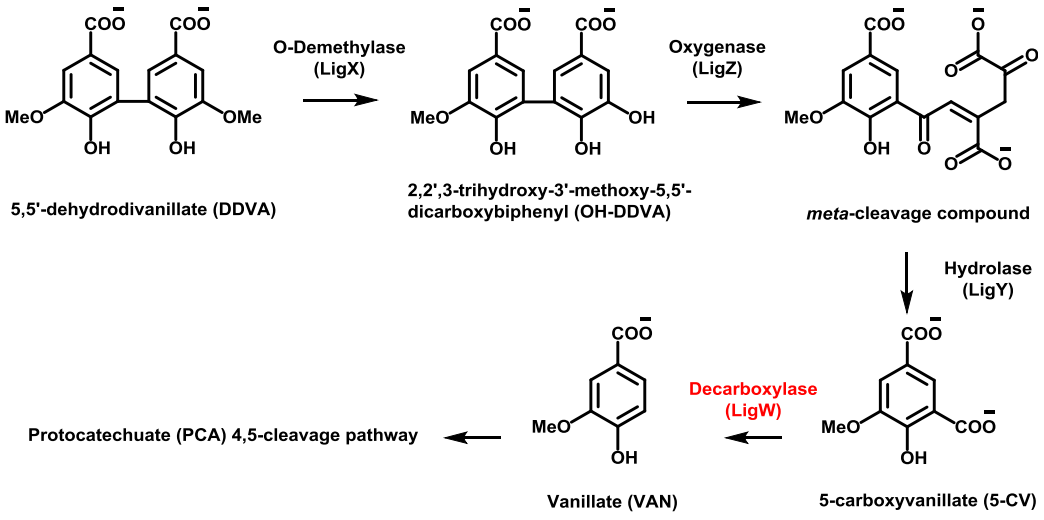


Figure 1.7: Sequence similarity network of COG2159 enzymes at BLAST E-value of 10^{-20} and 10^{-70}, respectively. LigW, γRSD, ACMSD, MSD, LigJ, LigY, and IDC are represented as squares in red, green, blue, yellow, black, purple, and pink, respectively.

LigJ, LigW, and LigJ are gene products of the lignin degrading bacteria *Sphingomonas paucimobilis* SYK-6 (*20, 21, 33*). This strain grows on various lignin related monoaryl and biaryl compounds as a sole carbon and energy source (*34*). Lignocellulose is the primary structural material of terrestrial plant cell walls and is comprised of the biopolymers cellulose, lignin, and hemi-cellulose (*35*). Lignin is an essential, highly breakdown resistant, aromatic heteropolymer composed of phenylpropanoids produced by the oxidative polymerization of monolignols (p-coumaryl, coniferyl, and sinapyl alcohols) (*36*). Various random ether and C-C bond linkages including, β-aryl ether, phenylcoumarane, biphenyl, pinoresinol, biaryl ether, and diaryl propane particularly hinder lignin decomposition (*37*). The β-ether linkage is most abundant in lignin; therefore, it serves as a rate limiting step to lignin decomposition, due to the lack of susceptibility to hydrolysis (*39*).

The degradation of lignin is cooperative in nature. It is initiated by various oxidoreductases secreted by white rot fungi and further degraded and mineralized by specialized bacterial enzymes. Processing of oxidized lignin fragments is important for the terrestrial life cycle; however, it is also of commercial and biotechnological interest, due to the vast availability of plant biomass for the production of renewable aromatic chemicals and biofuels (*35, 40*). Thus, there is a substantial need for understanding and deciphering microbial catabolic pathways that process oxidized lignin fragments in order to overcome the challenges in lignocellulosic refinement. Although, many of the genes responsible for degradation of lignin by SYK-6 have been identified, the three dimensional structures and catalytic reaction mechanisms for many of these enzymes remain to be elucidated (*41*). In the biphenyl degrader SYK-6, LigY and LigW catalyze

two consecutive reactions in the 5, 5'-dehydrodivanillate (DDVA) catabolic pathway and LigJ is found further downstream in the protocatechuate (PCA) 4,5-cleavage pathway (*41*, *42*). In SYK-6 biphenyl degradation shown in **Scheme 1.5**, DDVA is converted to 2,2',3-trihydroxy-3'-methoxy-5,5'-dicarboxybiphenyl (OH-DDVA) catalyzed by a non-heme iron-dependent demethylase LigX, followed by the formation of a meta-cleavage compound by extradiol dioxygenase LigZ, which is further hydrolyzed by the hydrolase LigY to 5-carboxyvanillate (5CVA) and decarboxylated by a nonoxidative decarboxylase LigW to vanillate. Further decomposition of vanillate occurs through the PCA 4, 5- cleavage pathway yielding the formation of pyruvate and oxaloacetate (*43*). 5-Carboxyvanillate decarboxylase (LigW) is found at the interface of two metabolic pathways in SYK-6, namely the biphenyl catabolism and the protocatechuate (PCA) 4, 5-cleavage pathway (*41*).



Scheme 1.5: Biphenyl degradation of lignin in *Sphingomonas paucimobilis* SYK-6.

Previous studies have indicated that LigW from SYK-6 is a 37 kDa non-oxidative decarboxylase, which functions independently of a metal or any other cofactor (*21*). Incubation of the cell extract containing LigW with a chelator also do not affect the activity. To test the substrate specificity of LigW, 3-methoxysalycilate (3MSA), 4-hydroxyisophthalate (4HIP), and isophthalate (IPA) were incubated with the enzyme. LigW was found to have a specific activity for 5-carboxyvanillate (5-CV); however, the activity was lowered in the presence of 4HIP and IPA by 51 and 79%. Furthermore, the decarboxylation reaction was carried out in D_2O and the resulted vanillate contained a deuterium at the C-5 of the substrate. Thus, presumably the reaction goes through a quinoid intermediate prior to decarboxylation. Since LigW is a non-oxidative decarboxylase, the carboxyl group is removed from the aromatic nucleus and replaced with a hydrogen that originates from solvent. This reaction is quite different than oxidative decarboxylation reactions, which replace the carboxyl group with a hydroxyl group (*44*).

A second LigW gene (LigW2) was identified, when the decarboxylation activity for 5-CV to VAN was not diminished, once the first LigW gene was disrupted by a gene replacement technique. The resulting mutant (DW) was able to perform the decarboxylation reaction with a specific activity of 120 ± 32 mU/mg, which is equivalent to 0.08 s^{-1} in presence of 5 mM DDVA. Furthermore, the disruption of LigW2 gene in DW mutant SYK-6, resulted in a double mutant (DDW). When the DW and DDW were tested for activity in presence of 5 mM DDVA, it was observed that the decarboxylation activity was diminished to 28% and 2.8% in comparison to the activity of the wild type cells, respectively. Thus it was determined that both genes contributed

to the overall decarboxylation of 5-CV; however, it seem as LigW2 plays a more important role in the degradation of 5-CV. Interestingly, LigW and LigW2 share 37% sequence identity and the closest homologue of the two proteins is γRSD from *Rhizobium radiobacter* with 33% and 28% sequence identity, respectively. LigW2 was also confirmed to perform the decarboxylation reaction in a non-oxidative manner by performing similar D_2O experiments as LigW (*45*).

Although the functions of LigW and LigW2 are known, the chemical mechanism, catalytic constants, metal and pH dependency, and crystal structures remain to be elucidated. It is interesting to understand how the active site of AHS members in COG2159 have evolved to perform hydration and decarboxylation reactions instead of the typical AHS hydrolysis. Active site analysis of γ-resorcylate decarboxylase (γRSD), α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD), and isoorotate decarboxylase (IDC) will aid in determining crucial catalytic residues of LigW and other COG2159 decarboxylases. It is likely that members of COG2159 share a similar structural topology, active site architecture, and metal ligands. Thus, by deciphering the mechanism of LigW, combined with information of characterized homologues, one may more accurately annotate uncharacterized enzymes in COG2159.

CHAPTER II

STRUCTURES AND SUBSTRATE DISTORTION OF 5-CARBOXYVANILLATE DECARBOXYLASE FROM *SPHINGOMONAS PAUCIMOBILIS* AND *NOVOSPHINGOBIUM AROMATICIVORANS*

Sphingomonas paucimobilis and *Novosphingobium aromaticivorans* are gram negative aerobic alphaprotobacteria found in soil, corals, and plant surfaces (*46*). These strains utilize various monoaryl and biaryl heterocyclic hydrocarbons as sole carbon and energy source (*47*). The oligotrophic nature of these sphingomonads allows for survival in nutrient deficient environments and formation of biofilms on pipes and reservoirs. The quick adaptation to man-made habitats allows these opportunistic pathogens to survive in tap and chlorinated waters (*48*). The prominent characteristic of these bacteria is also to degrade xenobiotic substances and become potentially hazardous, when introduced to humans (*49*). Although this family of bacteria is largely diversified, the antibiotic resistances and catabolic pathways by which various aromatic hydrocarbons are metabolized remain poorly understood (*50*). Additionally, these rod-shaped organisms are unusual, since the cellular membrane is composed of glycosphingolipids instead of lipopolysaccharides, which is common in eukaryotic organisms. This enables the bacteria to attract various hydrophobic compounds. Most of the research on aromatic hydrocarbon degradation has been done with several strains of sphingomonads according to published data (*52*).

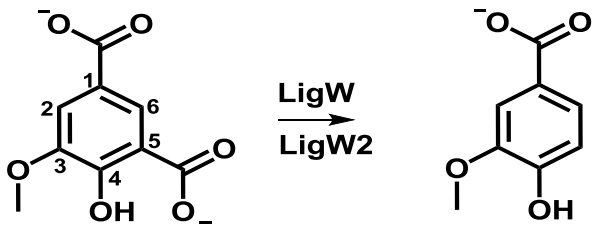
The *Sphingomonas* sp. SYK-6 isolated from kraft pulp effluent was shown to metabolize biphenyl and lignin related compounds; such as, 5, 5'-dehydrodivanillate, syringate, vanillate, and protocatechuate (*53*). In addition to degrading biphenyls,

Novosphingobium aromaticivorans F199 (DSM 12444) isolated from deep sediments has been found to degrade naphthalene, salicylate, xylene, and toluene (*52*). Infections of this strain have been shown to cause an autoimmune primary biliary cirrhosis in humans. This could be due to the ability of the bacteria to degrade alkyl-benzene compounds (*53*).

LigW (Accession number: AB033664) and LigW2 (Accession number: AB089690) of *Sphingomonas* sp. SYK-6 are members of COG2159 of the amidohydrolase superfamily (AHS) (*21, 45*). In general, enzymes from the amidohydrolase superfamily possess a mononuclear, binuclear, or trinuclear metal center embedded at the C-terminal end of a distorted $(\beta/\alpha)_8$-barrel (*55*). The SYK-6 LigW and LigW2 catalyze the non-oxidative C-C bond cleavage of 5-carboxyvanillate (5-CV) to vanillate (VAN) as illustrated in **Scheme 2.1**. Both genes are annotated as 5-carboxyvanillate decarboxylase and share 37% sequence identity. LigW2, currently annotated as amidohydrolase 2, of *Novosphingobium aromaticivorans* F199 (DSM 12444, locus tag: Saro_0799) shares the same function as the enzymes found in *Sphingomonas* sp. SYK-6 with 39% and 62% sequence identity with LigW and LigW2 in this strain, respectively. In the lignin related degrader SYK-6, these enzymes have demonstrated a vital role in the overall degradation of aromatic compounds carried out through the biphenyl catabolism and the protocatechuate 4, 5-cleavage pathway (*41*).

Multiple high resolution crystal structures of a Mn^{2+} substituted LigW and LigW2 from *Sphingomonas* sp. SYK-6 and *Novosphingobium aromaticivorans* DSM12444 have been solved. Crystal structures of LigW in complexed with vanillate (VAN), 3-methoxy-5-carboxybenzoate (MCB), and 5-nitrovanillate (5-NV) have been

resolved to 1.4 Å, 1.6 Å, and 1.7 Å, respectively. Furthermore, LigW2 and LigW2 D314N mutant structures in the presence and absence of 5-NV have been reported to a resolution of 1.1 Å and 1.8 Å, respectively. The inhibitor 5-NV is an unreactive structural homologue of 5-CV. Comparison of the individual crystal structures of 5-NV and 5-CV reveal that the two compounds are planer and share similar geometry. However, significant deviations from planarity are seen in the inhibitor bound structures. The crystallographic data is compared to the available crystal structures of α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD), γ-resorcylate decarboxylase (γRSD), and uracil-5-carboxylate decarboxylase (IDC) in hope to gain an insight of the COG2159 decarboxylation reactions.



Scheme 2.1: Reaction catalyzed by LigW from *Sphingomonas* sp. SYK-6 and LigW2 from *Novosphinogobium aromaticivorans* DSM1244.

MATERIALS AND METHODS

Materials. All chemicals, buffers, and purification reagents were purchased from Sigma-Aldrich unless otherwise stated. 5-nitrovanillate (5-NV) was synthesized according to published procedures (*56*). Pfu Turbo DNA polymerase and the Escherichia coli strains BL21 (DE3) and XL1-blue cells were obtained from Stratagene. The restriction enzyme DpnI was purchased from New England BioLabs and oligonucleotides were obtained from Integrated Data Technology through the Gene Technology Laboratory at Texas A&M University. Kanamycin, isopropyl β-D-thiogalactopyranoside (IPTG), and LB broth were acquired from Research Products International Corp. Protamine sulfate, Wizard Plus SV Miniprep DNA purification system, and Vivaspin centrifugal concentrator (10 000 MWCO) were obtained from MP Biomedicals LLC., Promega, and Fisher Scientific, respectively. Chromatographic gel filtration columns were purchased from GE Healthcare.

Expression and Purification of LigW from *S. paucimolbilis* and LigW2 from *N. aromaticivorans*. The recombinant plasmid SpCD00297569 (NYSGXRC clone ID: 9453b1BCt3p1) bearing the gene for LigW (gi|19110430) from *S. paucimobilis* was obtained through the PSI Material Repository (dnasu.asu.org). The NYSGXRC clone encoded codon-optimized LigW with an N-terminal Met-Ser-Leu insertion and a C-terminal Glu-Gly-His_6. The nucleotides coding for the His-tag were removed from the plasmid using standard site directed mutagenesis techniques with the complementary primer pair 5'- CGCATCACCTAATGACATCATCAC CATCACCACTG-3' and 5'-CAGTGGTGATGGTGATGATGTCATTAGGTGATGCG-3'. The QuikChange reaction product was transformed into E. coli XL1-blue cells and the plasmid was

isolated using the Wizard Plus SV Miniprep DNA purification system. After verifying the DNA sequence, the expression plasmid was transformed into E. coli BL21(DE3) cells by electroporation and then grown on LB-agar plates containing 50 µg/mL kanamycin for 14 hours at 37 °C. A single colony was grown overnight in 7 mL of LB medium containing 50 µg/mL kanamycin at 37 °C. The cell culture was used to inoculate 1 L of the same medium, which was further allowed to incubate by agitation at 37 °C. The temperature was lowered to 25 °C after the optical density at 600 nm reached 0.6, at which point protein expression was induced with 0.5 mM IPTG in the presence of 1.0 mM $MnCl_2$. The cells were harvested by centrifugation (12,000 x g at 4 °C for 10 minutes) after 14 hours of incubation.

The cell pellet was resuspended in 20 mM HEPES at pH 7.3 and then lysed by sonication using 5 second pulses for 20 minutes at 0 °C in the presence of 0.2 mg/mL phenylmethanesulfonyl fluoride (PMSF). The cellular debris was separated by centrifugation and the supernatant fluid treated with 10 mL 2% (w/v) protamine sulfate for 30 minutes to precipitate the nucleic acids. The protein was precipitated with 45% saturation of solid ammonium sulfate, resuspended in a minimal amount of 20 mM HEPES (pH 7.3), filtered with a 0.2 µm cellulose acetate syringe filter, and loaded onto a Hiload 26/60 Superdex 200 prep grade gel filtration column (Amersham Biosciences). The purity of the protein was confirmed by SDS-PAGE. The fractions were pooled and concentrated to 17 mg/mL using a Vivaspin centrifugal concentrator (10 000 MWCO) at 4500 rpm. A pNIC28-Bsa4 plasmid containing the LigW2 gene from *N. aromaticivorans* (UNP ID Q2GA79, gene Saro_0799, EFI Target 505250) was obtained from the laboratory of Dr. Steve Almo at the Albert Einstein School of Medicine through

the Enzyme Function Initiative (EFI). The expression and purification of LigW2 was accomplished under the same conditions as described for LigW from *S. paucimobilis*.

Construction and Purification of LigW2 D314N Mutant. The D314N mutant was constructed by subjecting the plasmid encoding LigW2 to the standard QuikChange site-directed mutagenesis protocols with the complementary primer pair: 5'-C GTT ATG TAC GCG ATG AAC TAC CCC TA C CAG TAC CAG TAC G-3' and 5'-C GTA CTG GTA GGG GTA GTT CAT CGC GTA CAT AAC G-3'. The DNA was transformed into E. coli XL1-blue competent cells and grown overnight on agar plates containing 50 μg/mL kanamycin. Single colonies were collected, and then grown overnight in LB-broth containing 50 μg/mL kanamycin. The D314N mutant was expressed and purified using the same protocol as the wild type enzyme. The purified protein was treated with ProTEV Plus protease (Promega) using the recommended protocol to remove the N-terminal His_6 tag. The ProTEV Plus protease was separated from the cleavage reaction by utilizing the HQ tag located at the N-terminus of the protein using HisTrap HP affinity column.

Crystallization of LigW from *S. paucimobilis* and LigW2 from *N. aromaticivorans*. The high resolution crystal structures of LigW from *S. paucimobilis*, LigW2 from *N. aromaticivorans*, the free structures of 5-CV and 5-NV were resolved in collaboration with Dr. Steven C. Almo laboratory at the Department of Biochemistry, Albert Einstein College of Medicine, at Bronx, New York. The PDB codes for Mn^{2+} substituted LigW are 4ICM for the apo form, 4L6D, 4NI8, and 4NG3 for the enzyme complexed with vanillic acid, 5-methoxyisophtalic acid, and 4-hydroxy-3-methoxy-5-nitrobenzoic acid, respectively. The crystal structures for LigW2 (EFI Target: 505250)

were deposited under the following PDB codes: 4INF and 4QTG with bound Ca^{2+} and Mn^{2+}, respectively, 4QRN in presence of Mn^{2+} and 4-hydroxy-3-methoxy-5-nitrobenzoic acid, 4QS6 corresponds to LigW2 D314N mutant with 4-hydroxy-3-methoxy-5-nitrobenzoic acid and no metal, and 4QS5 represents the LigW2 D314N mutant in presence of Mn^{2+} and 4-hydroxy-3-methoxy-5-nitrobenzoic acid.

Molecular Weight Determinations. The oligomeric states of LigW and LigW2 were determined using a 10/300 GL SuperdexTM 200 column from GE Healthcare calibrated with gel filtration standards (Bio-Rad Laboratories, Inc.) containing a lyophilized mixture of thyroglobulin (MW 670,000), bovine γ-globulin (MW 158,000), chicken ovalbumin (MW 44,000), equine myoglobin (MW 17,000), and vitamin B12 (MW 1,350) in 50 mM HEPES, pH 7.0. Various concentrations of LigW ranging from 0.5 mg/mL to 2.0 mg/mL were utilized at a flow rate of 0.5 mL/min with a pressure of 1.5 MPa.

RESULTS

Three-Dimensional Crystal Structure of LigW and LigW2. The crystal structure of LigW (334 amino acids) and LigW2 (351 amino acids) with Mn^{2+} in the active site were determined to resolution of 1.8 Å (PDB id: 4ICM) and 1.6 Å (PDB id: 4QTG), respectively. The ribbon representation of both proteins is provided in **Figure 2.1,** indicating Mn^{2+}•LigW as an octamer (panel **A**) and Mn^{2+}•LigW2 as a dimer (panel **B**). Each protein subunit consists of a central distorted $(\beta/\alpha)_8$-barrel domain, where the active site and divalent cation are localized at the C-terminal end of the β-barrel. An insertion domain is observed in these structures consisting of three helices α1', α2', and α3' (residues 11-57 in LigW and 23-70 in LigW2) between β-strand 1 and α-helix 1.

Additional helices are found before and after α-helix 5 (α4' and α5') and after α-helices 6 (α6') and 8 (α7'). Two 3_{10} helices, η1' and η2', appear after β-strand 6 and α-helix 7, respectively (**Figure 2.2A and B**).

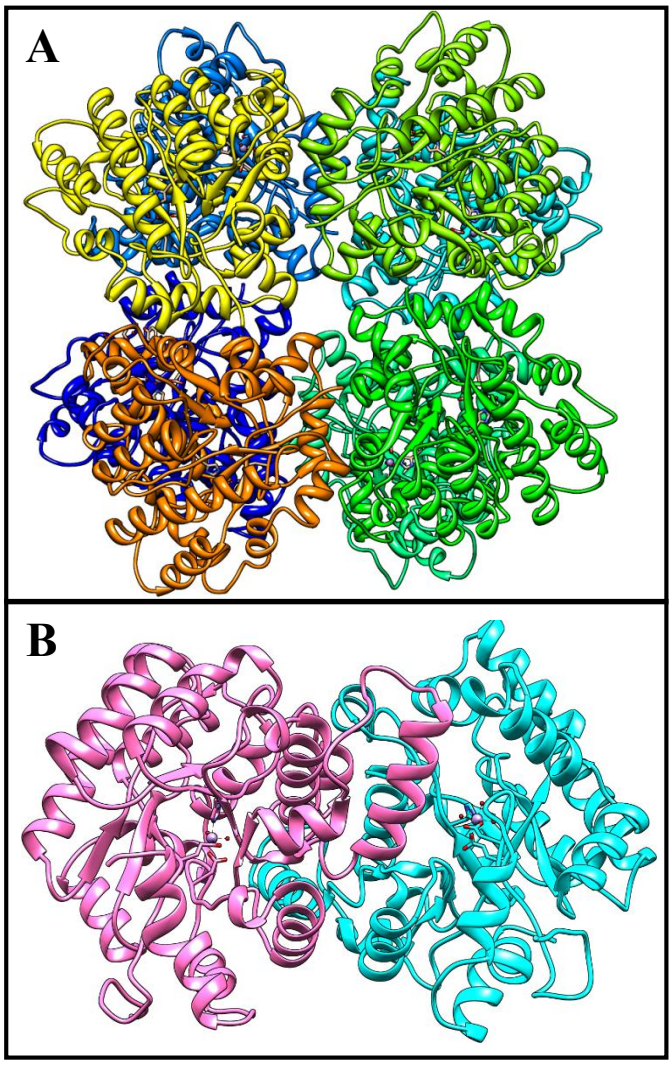


Figure 2.1: Ribbon representation for LigW with the eight subunits in the asymmetric unit distinguished by different solid colors (**A**). Ribbon representation of LigW dimer in pink and blue (**B**).

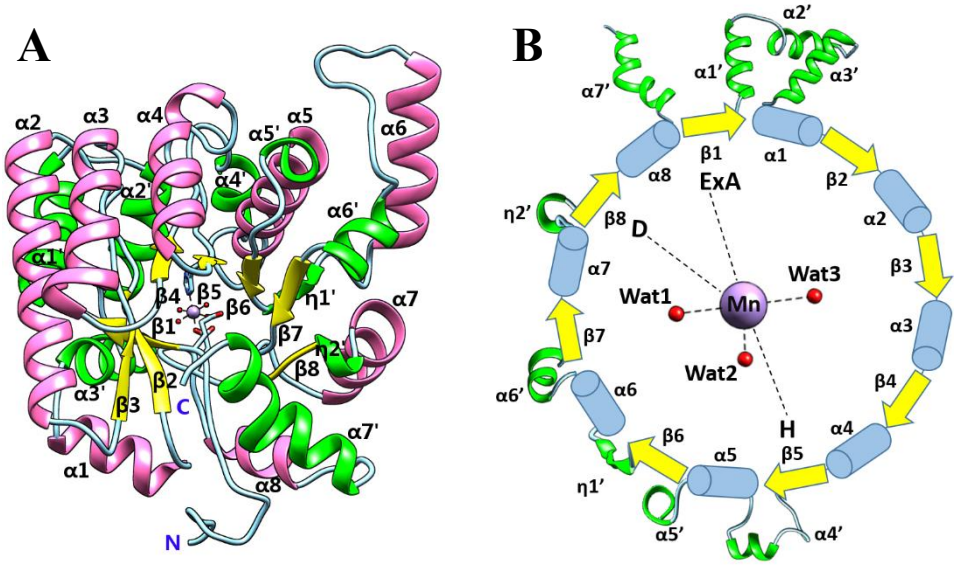


Figure 2.2: A ribbon representation of an individual monomer of LigW2 is depicted. The $(\alpha/\beta)_8$ barrel is shown with α-helices in pink and β-sheets in yellow. The insertion domain is colored in green and it includes α1'- α3' helices. The additional α-helices (α4'-α7') and 3_{10} helices (η1', η2') are also labeled and displayed in green. The Mn^{2+} ion is depicted as a purple sphere and it is clearly coordinated by 3 water molecules, glutamate from β-1, histidine from β-5, and aspartate from β-8. (**A**). Cartoon representation of the seccondary elements of LigW2 (**B**).

The α-helix loop region between β-sheets 7 and 8 from one subunit (residues 231-265 in LigW and 245-283 in LigW2) extends in the active site of the adjacent identical monomer forming a dimer with approximate interface surface area of 2872 $Å^2$ for the Mn^{2+}•LigW and 2690 $Å^2$ for the Mn^{2+}•LigW2. All protein monomers are nearly identical. Although LigW and LigW2 share 39% sequence identity, the overall active site architecture is conserved. The molecular weight of the purified LigW and LigW2 were determined to be ~ 1090 kDa and ~166 kDa, respectively by size exclusion chromatography, indicating that LigW is (monomer ≈ 38 kDa) likely a 32-mer and LigW2 (monomer ≈ 39 kDa) is a homotetremer in solution.

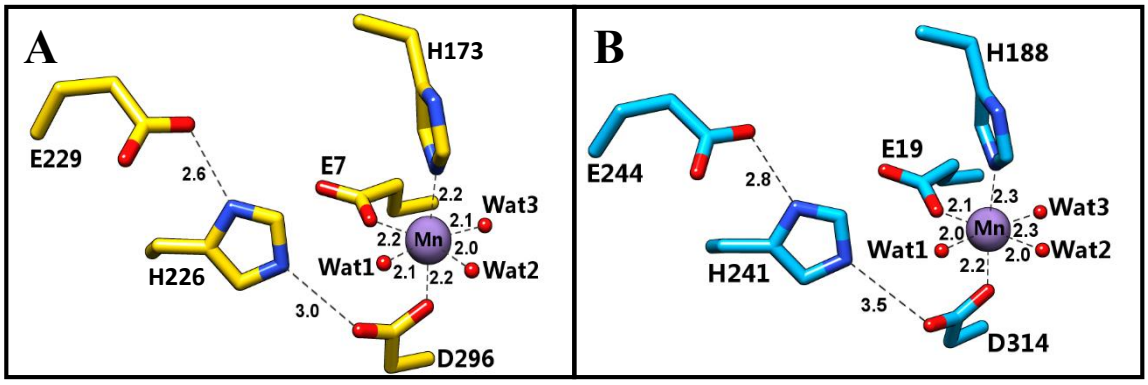
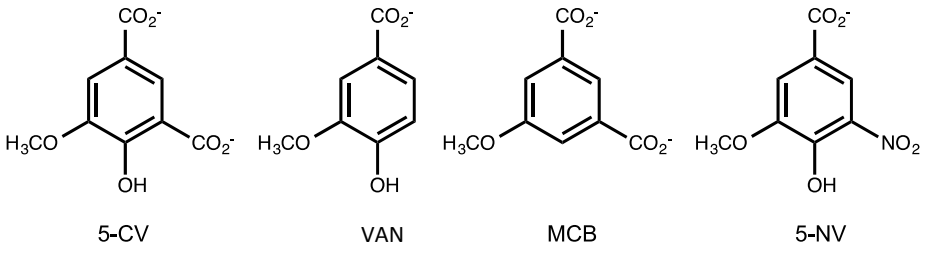


Figure 2.3: Active site architecture of LigW (**A**) and LigW2 (**B**). In the absence of added ligands, the manganese in the active site is coordinated to Glu-7, His-173, Asp-296 in LigW and Glu-19, His-188, and Asp-314 in LigW2 in addition to three water molecules.

Metal Binding Site. A single divalent cation is bound in the active site of LigW and LigW2. In LigW, the residues important for binding Mn^{2+} include a glutamate from β-strand 1 (Glu-7), a histidine from β-strand 5 (His-173), and an aspartate from β-strand 8 (Asp-296). These residues are highly conserved in LigW2 and include Glu-19, His-188, and Asp-314. In addition to these three residues, there are three bound water molecules that complete an octahedral coordination geometry. The conserved aspartate from β-strand 8 also forms a hydrogen-bond triad with conserved histidine (His-226 in LigW and His-241 in LigW2) and glutamate (Glu-229 in LigW and Glu-244 in LigW2) residues that originate from β-strand 6 in an HxxE motif. The structure of the metal binding site in LigW and LigW2 is illustrated in **Figure 2.3A and B**.

Structures of LigW Inhibitor Complexes. Crystal structures of LigW complexed with 3-methoxy-4-hydroxy-5-nitrobenzoate (5-NV), 3-methoxy-5-carboxybenzoate (MCB), and vanillate (VAN) were obtained at resolutions of 1.7 Å, 1.6 Å, and 1.4 Å, respectively. The structures of these compounds are presented in **Scheme 2.1**. In the structure of LigW with 5-NV, two of the metal-bound waters (Wat-1 and Wat-2 of **Figure 2.3A**) are displaced by one of the oxygens from the nitro group at C5



Scheme 2.2: Substrate 5-CV in comparison to the inhibitors VAN, MCB, and 5-NV.

and the hydroxyl group at C4, respectively. The octahedral geometry of the metal complex is retained similarly to LigW2 (**Figure 2.4A**). The nitro-substituent (a noncleavable carboxylate surrogate) in the LigW structure, further interacts with Arg-237 from the adjacent subunit. The carboxylate at C1 interacts electrostatically with Arg-45, and the methoxy group at C3 hydrogen bonds with the side chain hydroxyl from Thr-76. The phenyl group of the inhibitor is π-stacked with the side chain of Tyr-299 at a distance of 3.7 Å and it is further stabilized by a Tyr-197 from α6 helix of the same chain, which is perpendicular to the phenyl ring on the opposite side with a closest distance of 3.8 Å. In addition, Arg-45 originates from the loop that follows the α2' helix from the insertion domain. Tyr-299 is found within the loop that follows β-strand 8 and Thr-76 originates from the loop region after β-strand 2.

The crystal structures of LigW (PDB ids: 4L6D, 4NG3, and 4NI8) complexed with VAN, 5-NV, and MCB show that the substituent groups coordinated to the metal center are bent out of the ring plane, when bound within the active site. For example, the phenolic oxygen of VAN is distorted ~15^0 out the plane of the aromatic ring (**Figure 2.5A**). The carboxylate group of MCB and the nitro-substituent of 5-NV are also distorted from the plane of the aromatic ring with angles of ~8^0 and 11^0, respectively (**Figure 2.5B and C**). In addition, the angle between the phenolic group and the plane of the phenyl ring in the 5-NV ligand is ~ 22^0 (**Figure 2.5C**). These distortions are likely facilitated by the interactions with the bound Mn^{2+} ion. The structure of LigW with VAN contains a pentacoordinated metal center, where Wat1 is missing and Wat2 is replaced by the hydroxyl group of the substrate (**Figure 2.4B**). The structure of LigW complexed with MCB is essentially identical to that of the complex with 5-NV except

that the missing hydroxyl group from the inhibitor MCB is replaced by a water molecule that is coordinated to the divalent cation (**Figure 2.4C**). The rest of the molecular interactions with MCB are the same as found with 5-VN.

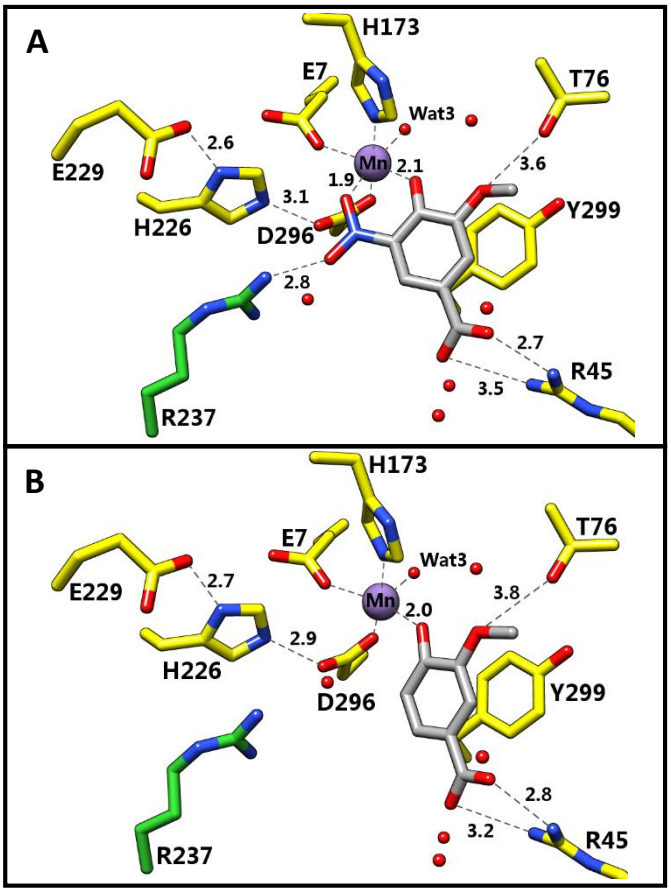


Figure 2.4: Active site architecture of LigW in the presence of 5-NV (**A**), VAN (**B**), and MCB (**C**) presented in gray-colored bonds. In all panels manganese is shown as a purple-colored sphere. The arginine residue (Arg-237) from the adjacent subunit is colored green. The coordinate and hydrogen bonds distance is shown in Å.

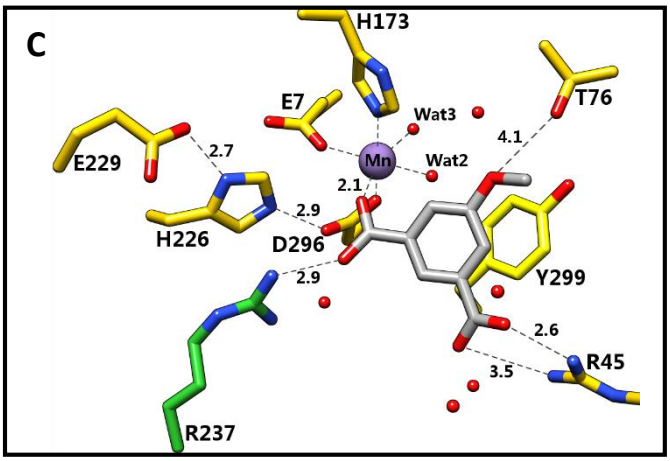


Figure 2.4: Continued.

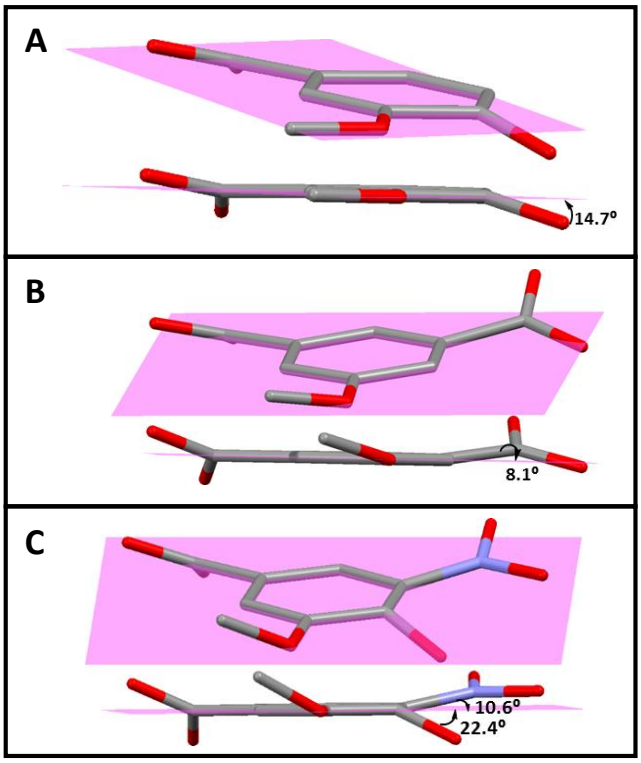


Figure 2.5: Geometric distortions of the bound ligands in the active site of LigW. The plane of the flat aromatic phenyl ring is represented in pink in each panel. The bound complex of the product vanillate (VNL) illustrates the bending of the phenolic oxygen at C4 at an angle of ~15° (**A**). The bound complex of the inhibitor 3-methoxy-5-carboxybenzoate (MCB) illustrates the bending of the bond to the carboxylate carbon at C5 at an angle of 8.1° (**B**). The bound complex of the inhibitor 3-methoxy-4-hydroxy-5-nitrobenzoate (5-NV) illustrates the bending of the phenolic oxygen at C4 at an angle of ~22° and the nitrogen of the nitro-group at C5 of ~11° (**C**).

Structures of LigW2 and D314N Mutant with Inhibitor Complexes. Crystal structure of wild type Mn^{2+}•LigW2 in presence of 5-NV were resolved to a high resolution of 1.07 Å (**Figure 2.1B**). The refined occupancy of Mn^{2+} in this structure was ~ 0.7 and the occupancy of 5-NV was ~ 0.9. The active site is located near the C-terminus of the $(\beta/\alpha)_8$-barrel domain and the substrate binding residues have been identified as: Tyr-51 from α2', Arg-58 from α3', Phe-212 from α5, His-241 from β6, Arg-252 from α6, and Tyr-317 from the loop region following α8 (secondary structure of LigW2 is illustrated in **Figure 2.2A and B**). As compared to LigW, the substrate binding opening is also located between α2', α3', and α5 from one subunit and α6 from the adjacent monomer. Upon binding of 5-NV into the active site of LigW2 (**Figure 2.6**), Wat1 and Wat2 are displaced by the ligand (**Figure 2.3B**), which coordinates to the metal center in a bidentate fashion retaining the octahedral geometry as observed in LigW·Mn^{2+}·5-NV complex in **Figure 2.4A**. The substrate and metal binding residues are identical to LigW with the exception of the residues coordinating to the C1 carboxyl group. In the LigW2, Tyr-51 from the α2' helix and Arg-58 from the α3' helix from the insertion domain interact with the C1 carboxyl group of the substrate analogue (**Figure 2.6 and 2.9A**). These residues are conserved in LigW and correspond to Tyr-39 and Arg-45; however, the orientation of the side chains of these residues differs significantly between the 5-NV bound LigW and LigW2 crystal structures (**Figure 2.4A and Figure 2.7**). Therefore, the Tyr-39 interaction with the C-1 carboxyl group is not observed in the structures of LigW with any of the inhibitors. Instead, Asn-244 is within a hydrogen bond distance with one of the oxygen atoms of the carboxyl at C-1 (**Figure 2.4**). The relative structures of LigW2 with Mn^{2+} complexed with 5-NV and LigW2 in presence of

Mn^{2+}, 4QRN and 4QTG in **Figure 2.10A**, are nearly identical with RMSD values of 0.15 Å between the atomic coordinates of the monomers from each structure. In addition, the conserved Arg-252 from the α6 helix in the neighboring monomer together with His-241 extend into the active site forming each 2.9 Å hydrogen bonds with the nitro group as shown in **Figure 2.9A**. The analysis of the electron density of the bound ligand suggest that the geometry is distorted from the main phenyl ring plane, composed of C1, C2, C6, and the carbon of the carboxylate at C1, by another plane formed by the hydroxyl group, C3, C4, and C5. Also, a third plane is formed by the nitrogen atoms of the nitro group and C-1 (**Figure 2.8**). The hydroxyl group and the nitro group of 5-NV

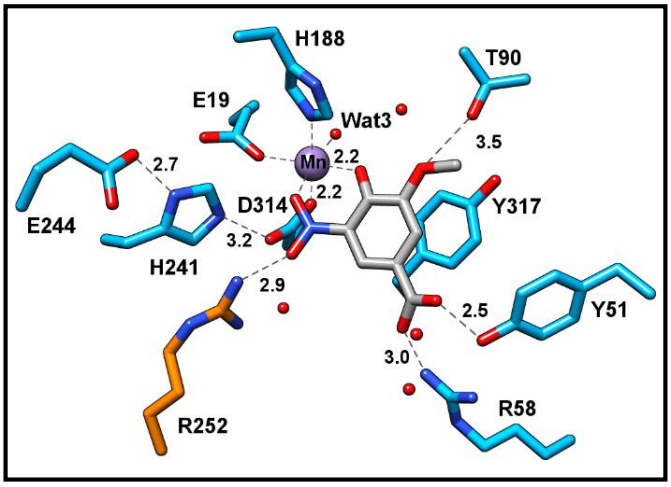


Figure 2.6: Active site architecture of LigW2 in the presence of 5-NV presented in gray-colored bonds. Manganese is shown as a purple-colored sphere. The arginine residue (Arg-252) from the adjacent subunit is colored orange. The coordinate and hydrogen bonds are shown as black dotted lines with distances in Å.

are ~13° and ~23° off the main benzene ring plane, respectively and the C-N bond is elongated to 1.50Å in comparison to the C-N bond of the crystalized free compound 1.44Å (**Figure 2.9B**). The arrangement of the substrate binding residues in the active

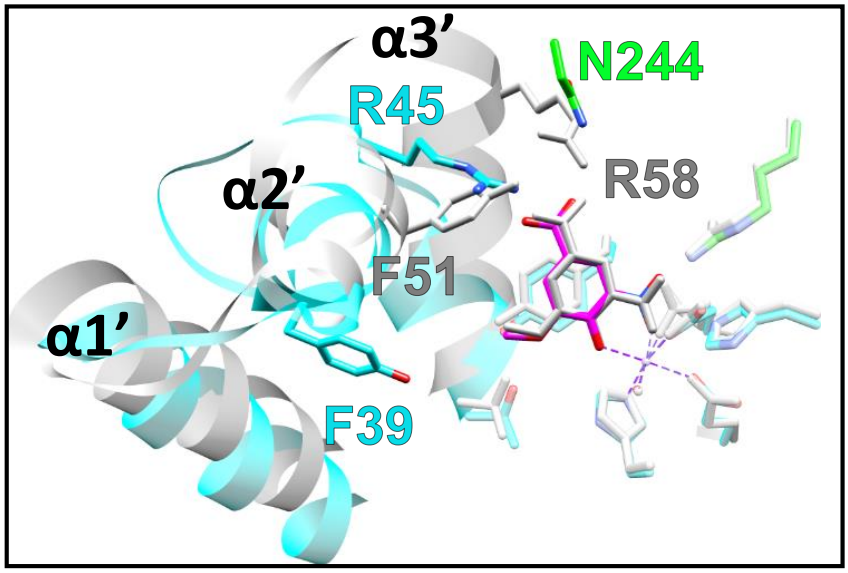


Figure 2.7: Superposition of wild-type LigW (PDB id: 4NG3, all atoms and ribbons are in cyan, 5-NV is in magenta, residues from the next subunit are in green, labels are colored accordingly for each residue, and oxygen and nitrogen atoms are represented in red and blue, respectively) and LigW2 with Mn^{2+} in presence of 5-NV (PDB id: 4QRN, all atoms and ribbons are represented in gray and all the labels are colored accordingly). The ribbons and hydrogen bond distances were omitted exception to the insertion domains in both enzymes, cyan in LigW and gray in LigW2.

site is nearly identical between the 5-NV bound Mn^{2+}· LigW2 and the apo Mn^{2+}· LigW2 structures and remain rigit with the exception of a small lateral movement of the insertion domain containg Tyr-51 and Arg-58, which form strong polar contacts with the C-1 carboxyl group. The phenyl ring of 5-NV is further sandwiched between Phe-212 and Tyr-217 as illustrated in **Figure 2.10A**, these hydrophobic residues are also seen in LigW and correspond to Tyr-197 and Tyr-299, respectively, shown in **Figure 2.4A** (Tyr-197 was omitted for clarity). Further, the methoxy group is in close proxymity to

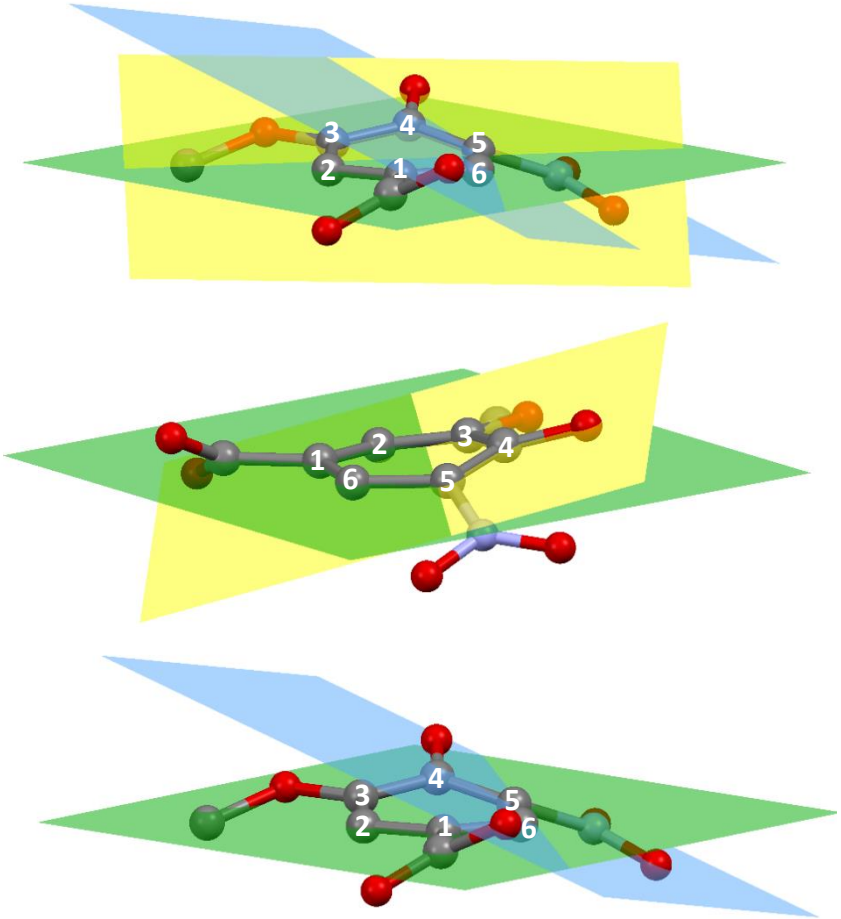


Figure 2.8: Geometric distortion of LigW2 bound 5-NV (PDB id: 4QRN).

hydrogen bond with Thr-90, which is also consrved in LigW and corresponds to Thr-76. In addtion, it is observed from the 1.07 Å crystal structure of LigW2 with the nitro inhibitor (PDB: 4QRN) that the arrangement of amino acids around the metal center remain the same in presence or absence of 5-NV with exception of Asp-314, which adopts two alternate conformations in presence of 5-NV. It is to be noted that the altenate conformation of Asp-314 colored in green in **Figure 2.9A** places the side chain 3.0 Å away from the C-5 of 5-NV; whereas the other conformation (colored in gray) uses the same oxygen atom to form a hydrogen bond with His-241.

The LigW2 D314N mutant high resolution crystal structures in presence of 5-NV (PDB id: 4QS6) and Mn^{2+}· 5-NV (PDB id: 4QS5) were solved to ~ 1.8 Å each. The active site of the mutant with Mn^{2+} and 5-NV is very similar to the wild-type enzyme in complex of the nitro-inhibitor (PDB id: 4QRN) shown in **Figure 2.10B**. The overall LigW2 conformation was not affected by the N314 mutation with an RMSD of ~ 0.3 Å for the superpositon of the individual monomers in 4QRN(wild type) with both 4QS5 and 4QS6 mutants. Interstingly, the *distorted* conformation of the nitro compound is still seen in both mutants with (**Figure 2.11A**) and without metal (**Figure 2.11B**) as compared to the rest of the 5-NV liganded structures. In the figure without Mn^{2+}, it is observed that the metal center is replaced by a loosely bound water molecule. The average RMSD value between the atomic coordinates of 5NV in both mutant structures is 0.1 Å. The superimposed mutants are shown in **Figure 2.11C** indicating that the rigidity of the active site is retained with small changes of Arg-58 between the two structures.

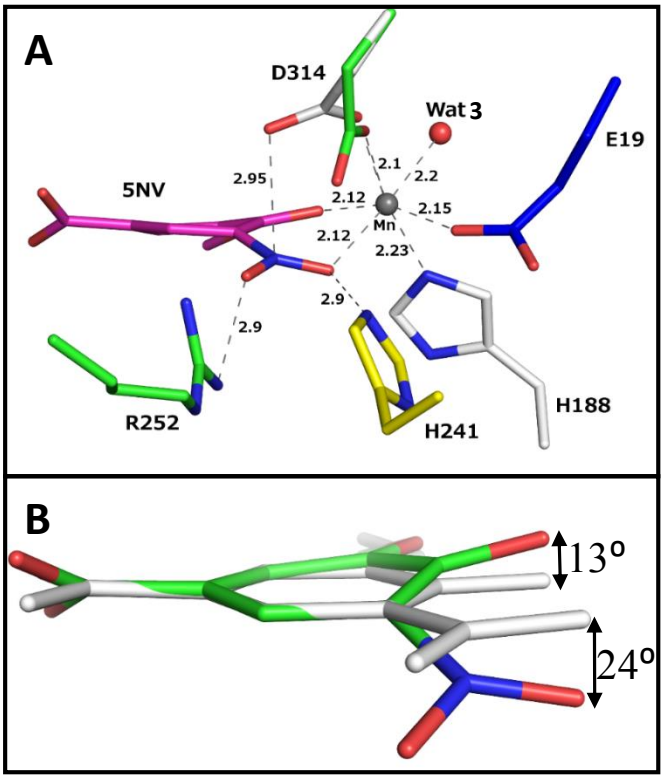


Figure 2.9: Active site of the 1.07 Å structure of LigW2 (PDB id: 4QRN) , Mn^{2+} is shown as a gray sphere, 5-NV is shown in magenta, Glu-19 from the AxE motif from β-1 is in yellow, His 188 from β-5 in gray, His-241 from β-6 in yellow, Arg-252 from the adjacent subunit in green, and the two alternate conformations of Asp-314 from β-8 in presence of 5-NV is depicted in gray and green. Hydrogen and coordination bonds are displayed as dotted lines and oxygen atoms are red and nitrogen atoms are blue (**A**). Geometric distortion of 5-NV in the active site of LigW2 (PDB id: 4QRN, Chain A) in green compared to the free solved 0.96Å structure of 5-NV in gray (**B**).

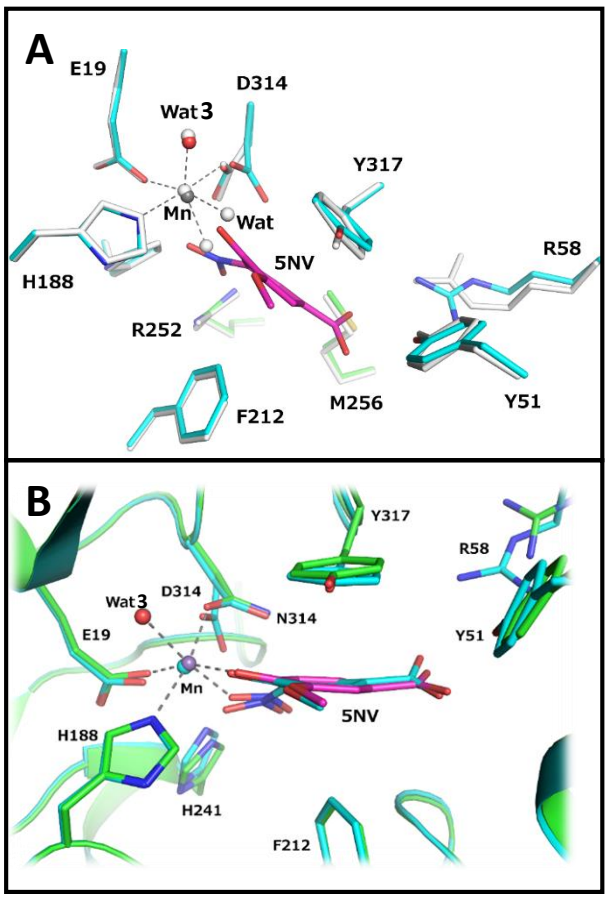


Figure 2.10: Superposition of wild-type LigW2 with Mn^{2+} in absence (PDB id: 4QTG) and presence of 5-NV (PDB id: 4QRN). The nitro inhibitor is shown in magenta and the amino acid side chains are in cyan (chain A) and green (Chain B) in the PDB id: 4QRN structure. Otherwise, nitrogen atoms are blue and oxygen atoms are red. All atoms in the Mn^{2+} LigW2 (PDB id: 4QTG) are shown in gray including the waters and metal center (**A**). Superposition of the wild-type LigW2 with Mn^{2+} and nitro inhibitor (PDB id: 4QRN, all atoms are in cyan) and LigW2 D314N mutant with Mn^{2+} and in presence of 5-NV (PDB id: 4QS5, all atoms are in green and 5-NV is colored in magenta). The nitrogen and oxygen atoms are colored in blue and red, respectively (**B**).

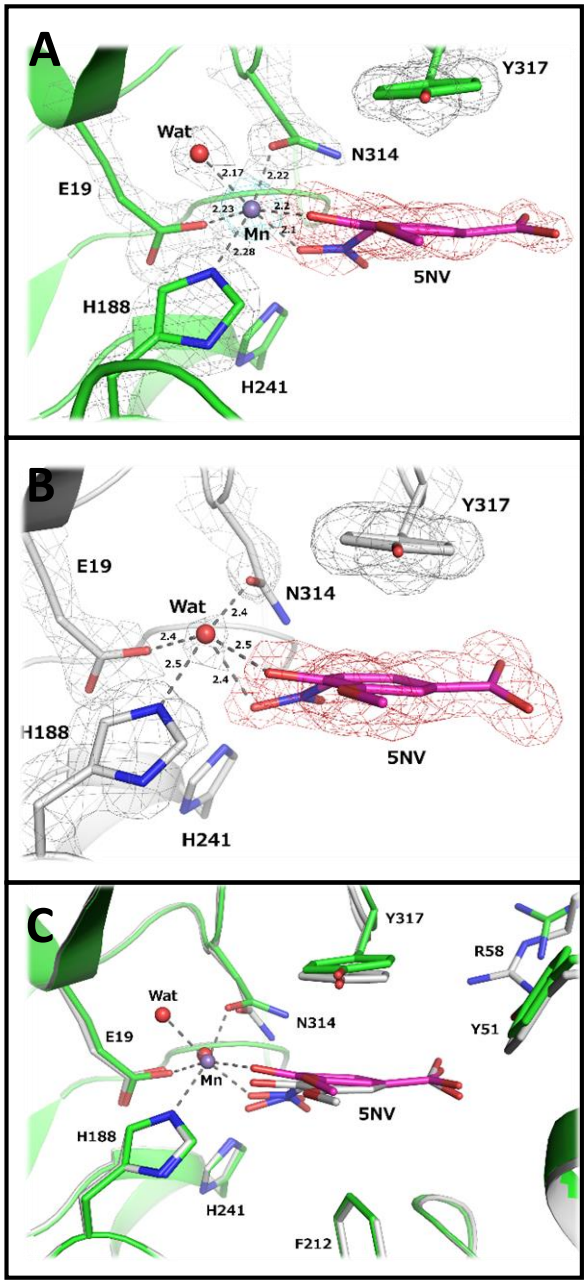


Figure 2.11: Distortion of 5-NV in the active site of LigW2 D314N mutant. Active site view of 4QS5 (**A**) and 4QS6 (**B**). Superposition of A and B with all atoms in 4QS6 colored in gray (**C**). Oxygen and nitrogen atoms are shown in red and blue.

DISCUSSION

Structure Elucidation of LigW and LigW2. The crystal structure of LigW from *S. paucimobilis* and LigW2 from *N. aromaticivorans* were determined to high resolution. The enzymes fold as a distorted $(\beta/\alpha)_8$-barrel and a single divalent cation is bound in the active site, which is located at the C-terminus of the barrel. Molecular weight determinations of LigW and LigW2 are most consistent with an association of 32 and 4 subunits in solution, respectively. Enzyme molecules form tight dimers in the crystal. The asymmetric unit in the LigW structures contain four similar dimers packed as an octamer and two monomers inside the dimer are connected by a non-crystallographic two-fold axis. Each monomer is nearly identical. In LigW, α6 helix formed from residues 235-245 of every monomer closes the active site cavity of the adjacent monomer within every dimer. The interactions between the dimers inside the octamers are much weaker than the interactions between the monomers inside the dimers. Similar observation are made in the structures of LigW2 and the minimal structural and functional unit is presumed to be a homo-dimer. Single subunits of LigW (334 residues) superimpose with its closest structural neighbors γ-RSD (PDB id: 2DVT), ACMSD (PDB id: 2HBV), and IDC (PDB id: 4HJW) with Cα atoms of 254, 171, and 205 residue pairs with RMSD values of 1.1 Å, 1.3 Å, and 1.1 Å, respectively. The overall topology of LigW and LigW2 are conserved within the three structural homologues and differ most noticeably from γ-RSD, ACMSD, and IDC because of an insertion domain located between β-strand 1 and α-helix 1.

Substrate and Metal Binding Residues. The Mn^{+2} cation is bound to three residues in both enzymes, namely E7, H173, and D296 in LigW and E19, H188, and

D314 in LigW2, together with three water molecules form a typical for this metal octahedral coordination. This array of active site metal binding ligands is similar, but not identical, to the other structurally characterized decarboxylases from COG2159. In ACMSD and IDC the single divalent cation is coordinated to four protein residues: two histidine residues from β-strand 1, a histidine from β-strand 5, and an aspartate from β-strand 8. The two histidine residues that coordinate the single divalent cation in the active site from β-strand 1 in ACMSD and IDC are found within an HxH motif that is very common among the structurally characterized members of the amidohydrolase superfamily (*23*). In contrast, the metal binding residues in γ-RSD include a glutamate and histidine from β-strand 1 (ExH motif) along with a histidine from β-strand 5 and an aspartate from β-strand 8. In LigW and LigW2, the conserved histidine in the HxH and ExH motifs from β-strand 1 of ACMSD, ICD, and γ-RSD is substituted with an alanine. In the three-dimensional structure of LigW and LigW2, the coordination site corresponding to this histidine residue is occupied by water. The functional perturbations that are imposed by these differences in the specific metal ligands contained within the active sites of LigW, LigW2, IDC, ACMSD, and γ-RSD are not clear.

The substrate binding residues in both proteins are inferred from the 5-NV bound structure. It is seen that the nitro inhibitor interacts with a highly conserved arginine residue from the adjacent monomer. This residue is conserved in all COG2159 characterized decarboxylases. Furthermore, the C-1 carboxyl group is stabilized by residues with different locations within the insertion domain of LigW and LigW2. The methoxy group is hydrogen bonded to a threonine residue and the phenyl ring is

sandwiched between two hydrophobic amino acids seen in both proteins. The active site architecture of the bound and unbound 5-NV is virtually identical with small conformational changes seen from the residues residing within the insertion domain. There are two conformations of the highly conserved aspartate from β-strand 8 with one placing one of the side chain oxygen atoms close to the C-5 of the phenyl ring. Otherwise, the same oxygen is hydrogen bonded with the histidine residue from β-strand 6.

Relevance of the Enzyme Bound 5-NV Distortion. The nitro inhibitor 5-NV and the substrate 5-CV were crystalized as free compounds and both have nearly the same planer geometry. However, the inhibitor exhibits a *distorted* conformation in the active site of LigW2 with significant deviations from planarity and thus perhaps resembles a probable intermediate. The nitro group is tilted ~ 10^0 and $23°$ off the benzene ring in both LigW and LigW2 inhibitor structures. However, the value of the C-2-C1-C6-C5 torsion angle of ~ 9.5^0 seen in PDB id: 4QRN with the C-5 puckered out of the plane is only seen in the LigW2 wild-type and D314N mutant structures. The LigW2 structures in presence and absence of the inhibitor show that the active site architecture is rigid and the binding cavity accommodates the distorted conformation of the bound ligand.

Comparison of the two D314N structures, one with 5-NV·Mn^{2+} complex and 5-NV-only complex has confirmed the notion that the ligand binding is not affected by the presence of the metal center. The 1.07 Å structure (PDB id: 4QRN) of LigW2 shows that the nitro group is bent away from the benzene ring; however, the hydroxyl group remains in plane with C3-C4-C5, which establish a different plane and bends the phenyl

ring. LigW2 crystal structures clearly articulates how the substrate binds into the active site and how presumably the substrate is converted to a transition state. Comparison of the crystal structures of the nearly planer free compounds, 5-NV and 5-CV, with the structure of the *distorted* enzyme bound 5-NV ligand, suggests that the diviation of planarity is mediated by the active site of the enzyme and interactions with the metal center. Therefore, perhaps the enzyme bound 5-NV closely resembles the geometry and charge distribution of a possible intermediate. The comparison of the two 5-NV bound molecules in LigW and LigW2, respectively indicates that the LigW bound 5-NV is planar with the nitro group and hydroxyl group deviating from the phenyl ring by and ~11° and ~22°; whereas, in the 5-NV in LigW2 the hydroxyl group forms a plane with C3-C4-C5 "splitting" the ring into two separate planes. This observation could be due to structural refinements; thus, it is believed that the 1.07 Å structure closely resembles the actual geometry and binding mode of the nitro inhibitor, which is mediated by the tight fit of the analogue in the active site and the coordination of the nitro group and hydroxyl group to the metal center.

Structural Comparison of LigW and γ-RSD. γ-RSD is the closest structurally characterized homologue of LigW with an overall sequence identity of 33%. This enzyme also catalyzes the decarboxylation of an ortho-substituted hydroxybenzoic acid. In an attempt to identify the functional roles of the catalytic residues in LigW and the other decarboxylases from COG2159, the crystal structure of the substrate-bound complex of γ-RSD (**Figure 2.12B**, PDB id: 2DVU) is compared with the structure of LigW in the presence of the nitro-analogue inhibitor (**Figure 2.12A**). The structure of γ-RSD, complexed with γ-resorcylate, reveals that the divalent cation is ligated in a

fashion similar to that of LigW, but with two notable exceptions. The ligand corresponding to His-10 in γ-RSD is replaced with a water molecule in LigW. In γ-RSD, the lone carboxylate group of the substrate makes a bidentate interaction with the active site metal ion, whereas in LigW, only one of the two oxygen atoms in the corresponding nitro group coordinate to the metal ion. In γ-RSD, neither of the two phenolic oxygen atoms interacts directly with the divalent cation, whereas in LigW the phenolic oxygen is ligated directly to the metal ion.

The difference in the observed binding geometry of these ligands by these two enzymes may be due to the substitution of a nitro group for the carboxylate in the inhibitor 5-NV. Alternatively, the structure of γ-RSD with γ-resorcylate may be an unproductive complex since this substrate is not decarboxylated by the crystalline enzyme. Evidence in support of the notion that the LigW structure represents a more realistic view of the enzyme-substrate complex can be found in the structure of LigW with the product vanillate (**Figure 2.4B**). In this structure, the hydroxyl group at C4 interacts directly with the metal ion in the active site. This conclusion is further supported by the observation that in the structure of LigW with 3-methoxy-5-carboxybenzoate (**Figure 2.4C**), the carboxylate of the inhibitor interacts with the metal ion via a single oxygen.

Comparison of LigW and IDC. The high-resolution crystal structure of the zinc-dependent 5-carboxyuracil decarboxylase (IDC) was determined previously in the presence of 5-nitrouracil as presented in **Figure 2.12C** (PDB id: 4HK6). The structure of this complex is very similar to that of the nitro-analogue inhibitor (5-NV) bound in the active site of LigW. In IDC, the active site zinc is coordinated to the nitro group and

the adjacent oxygen of the bound inhibitor. In addition, Arg-262* from the adjacent subunit interacts with the nitro-substituent of the inhibitor in a manner identical to that of Arg-237* in LigW (**Figure 2.4A**). This residue is highly conserved among all COG2159 decarboxylases. The only difference in the metal-binding ligand set is the coordination of His-14 in IDC and a water molecule in LigW. However, it should also be noted that there are no reported distortions from planarity of the nitro group and the oxygen attached to C4 of the inhibitor in the IDC structure.

Comparison of LigW and ACMSD. The structure of ACMSD has been determined in the presence of the inhibitor dihydroxyacetone phosphate (PDB id: 2WM1). Unfortunately, the structure of this enzyme has not been determined in the presence of an inhibitor or other ligand that more resembles the substrate or the decarboxylated product. However, one can utilize the information that has been obtained from the structures of LigW and IDC to postulate how the substrate may bind productively in the active site of ACMSD. In this model of ACSMD (**Figure 2.12D**) the carboxylate to be cleaved from C3 and the amino group at C2 are proposed to associate with the divalent metal ion in a manner similar to the orientation of the nitro-substituent and the hydroxyl group of the inhibitor with LigW (**Figure 2.12A**) and ICD (**Figure 2.12C**). The carboxylate at C3 is predicted to interact electrostatically with the arginine residue of the adjacent subunit in ACMSD (Arg-239*). The C1-carboxylate will likely ion-pair with Arg-47 in ACMSD similar to the interaction that is found in LigW with Arg-45 and ICD with Arg-68. It has previously been suggested for ACMSD that the carboxylate at C3 will coordinate with Arg-51 and that the C1-carboxylate will

coordinate with Arg-239* (24). However, this orientation is inconsistent with the structures of the inhibitors bound in the active site of IDC and LigW.

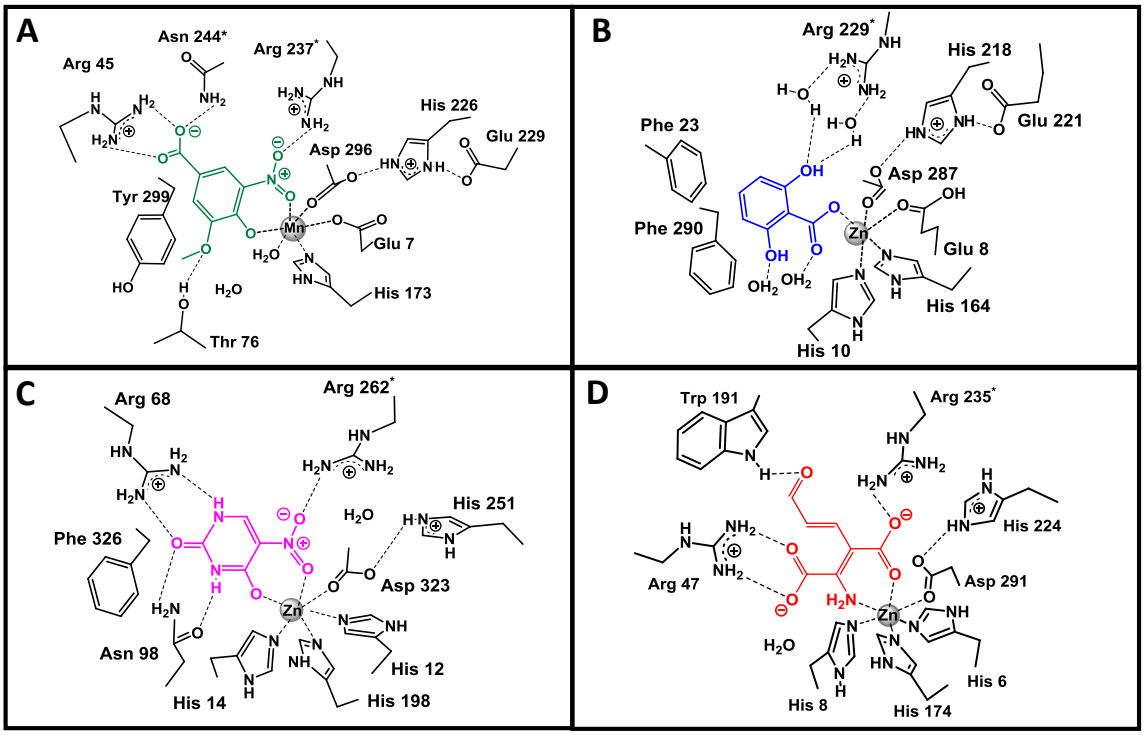


Figure 2.12: Active sites structures of representative decarboxylases from COG2159. (**A**) LigW complexed with 5-NV. (**B**) γ-RSD complexed with resorcylate (PDB id: 2DVU). (**C**) IDC complexed with 5-nitrouracil (PDB id: 4HK6). (**D**) Proposed model for the binding of α-aminomuconate-ε-semialdehyde in the active site of ACMSD.

Implications for Catalytic Mechanism of Decarboxylation. There are a number of plausible reaction mechanisms for the decarboxylation of substrates by LigW. In the reaction catalyzed by LigW, the carboxylate at C5 must be replaced by a proton. The protonation of the C5 carbon may occur after the decarboxylation event with a carbanion intermediate, or C5 may be protonated prior to decarboxylation via delocalization of the electrons from the hydroxyl group at C4. The initial protonation of C5 could also occur via hydration of the double bond between C4 and C5. Decarboxylation would then proceed via expulsion of the newly added hydroxyl group at C4. Since the decarboxylases of COG2159 have apparently evolved from within an enzyme superfamily that catalyzes the hydrolysis of carboxylate and phosphate esters, it is not unreasonable to imagine the activation of a water molecule for nucleophilic attack in the active site of LigW. However, the crystal structures of LigW and LigW2, determined in the presence of substrate analogues, do not support the activation of a water molecule for the hydration of the substrate. The chemical mechanism of the decarboxylation reaction will be addressed in more detail in the chapter that follows.

CHAPTER III

CATALYTIC MECHANISM OF 5-CARBOXYVANILLATE DECARBOXYLASE (LIGW) FROM *SPINGOMONAS PAUCIMOBILIS* SKY-6

5-Carboxyvanillate decarboxylase (LigW) catalyzes the conversion of 5-carboxyvanillate (5-CV) to vanillate (VAN). This transformation is at the interface of biphenyl catabolism and the protocatechuate (PCA) 4, 5-cleavage pathway in the degradation of lignin (*21*). Lignin is the most abundant heterogeneous aromatic polymer contained in the structural material of plant cell walls (*57*). The degradation of lignin is initiated by various oxidoreductases secreted by white rot fungi and further mineralized by specialized bacterial enzymes. Processing of oxidized lignin fragments is important for the terrestrial life cycle, and it is also of commercial and biotechnological interest (*58*). The high resolution crystal structures of two enzymes sharing the same 5-carboxyvanillate decarboxylase function, namely LigW from *Sphingomonas Paucimobilis* SYK-6 and LigW2 *Novosphingobium aromaticivorans* DSM 12444, have been previously resolved in the apo form and in presence of inhibitors vanillate (VAN), 3-methoxy-5-carboxybenzoate (MCB), and 5-nitrovanillate (5-NV). Sequence alignments demonstrate that LigY, LigW, and LigJ from *Sphingomonas Paucimobilis* SYK-6 belong to COG2159 from the amidohydrolase superfamily (AHS) of enzymes. The AHS is composed predominately of hydrolytic enzymes that catalyze the cleavage of ester and amide bonds at carboxylate or phosphate centers contained within amino acids, sugars, and nucleic acid containing substrates (*59*). Proteins within the AHS fold

as a distorted $(\beta/\alpha)_8$-barrel and most of the structurally characterized enzymes within this superfamily contain 1-3 divalent cations bound within the active site.

Proteins from COG2159 are the only known enzymes in the AHS that catalyze decarboxylation and hydration reactions. As previously noted, there are seven known function in COG2159: γ-resorcylate decarboxylase (γRSD) (*60*), α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) (*61*), OH-DDVA *meta*-cleavage compound hydrolase (LigY) (*19*), 4-oxalomesaconate hydratase (LigJ) (*20*), 5-carboxyvanillate decarboxylase (LigW) and (LigW2) (*21*), uracil-5-carboxylate decarboxylase (IDC) (*22*), and 6-methylsalicylic acid decarboxylase (MSD) (*23*). The crystal structures of γ-RSD (*24*), ACMSD (*25*), and IDC (*22*) are known. All of these enzymes contain a single divalent cation in the active site.

A sequence similarity network (SSN) of COG2159 enables the visualization of the phylogenetic relationship amongst the decarboxylases in this superfamily (*62*). The sequence similarity network in Figure 3.1 was constructed at a BLAST *E*-value cutoff of 10^{-70} and groups of related sequences have been arbitrarily numbered. At this stringency level, Group 1 contains LigW, LigW2 and RSD, whereas Groups 2, 17, 26, 30 and 31 contain ACSMD and LigJ, MSD, LigY, and IDC, respectively. γ-RSD is the closest structurally characterized homologue of LigW from *Sphingomonas Paucimobilis* SYK-6 with ~33% sequence identity, and this enzyme similarly catalyzes the decarboxylation of a substituted hydroxybenzoate substrate. Here we propose a chemical reaction mechanism for the decarboxylation of 5-carboxyvanillate by LigW based on the three-dimensional X-ray structures and the catalytic properties of this enzyme determined as a

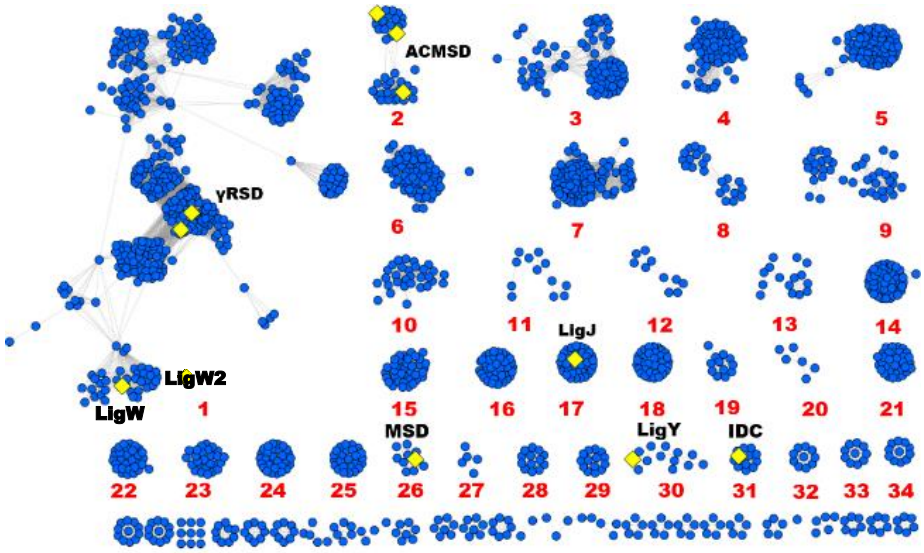


Figure 3.1: Sequence similarity network (SSN) of COG2159 at a BLAST E-value of 10^{-70}. At this stringency, the sequences represented by nodes in blue are separated into discrete clusters containing similar sequences connected by edges in gray. The length of the edges represents the degree of pairwise sequence relationship between clustered sequences; highly coalesced clusters share higher sequence similarity than nodes with fewer and longer connections. The known catalytic functions for enzymes in COG2159 are represented by yellow diamonds. Group 1 contains several subgroups of enzymes, including LigW from *S. paucimobilis* (PDB id: 4ICM), *N. aromaticivorans* (PDB id: 4QTG) and γRSD from *Rhizobium* sp. (PDB id: 2DVU) and *Polaromonas* sp. (PDB id: 3S4T). Group 2 contains ACMSD from human (PDB id: 2WM1) and ACMSD from *P. fluorescens* (PDB id: 2HBV). Group 17 includes LigJ from *Rho. palustris* (PDB id: 2GWG). Group 26 contains MSD from *A. clavatus* (gi|401829650) and Group 30 includes a homologue of LigY from *R. eutropha* (gi|73538619). IDC from *M. anisopliae* (PDB id: 4HJW) belongs to Group 31.

function of isotope effects and changes in the magnitude of the kinetic constants by pH and mutation of active site residues.

MATERIALS AND METHODS

Materials. 5-Carboxyvanillate was purchased from AApin Chemicals Limited, vanillate and 3-methoxy-5-carboxybenzoate (MCB) were obtained from Sigma, and 5-nitrovanillate (5-NV) was synthesized as described previously (*56*). *Pfu Turbo* DNA polymerase and E. coli strains BL21(DE3) and XL1-blue were purchased from Stratagene. The DpnI restriction enzyme was acquired from New England BioLabs and all DNA primers were obtained from Integrated Data Technology through the Gene Technology Laboratory at Texas A&M University. Kanamycin, isopropyl β-D-thiogalactopyranoside (IPTG), and LB broth were acquired from Research Products International Corp. Deuterium oxide (D_2O) was purchased from Cambridge Isotope Laboratories. Protamine sulfate, Wizard Plus SV Miniprep DNA purification system, Spectra/Por Molecularporous Membrane Tubing (3500 MWCO), and Vivaspin centrifugal concentrator (10 000 MWCO) were obtained from MP Biomedicals, Promega, VWR Scientific, and Fisher Scientific, respectively. The HisTrap HP column was purchased from GE Healthcare. The isolation of LigW from *S. paucimobilis* SYK-6 and LigW2 of *N. aromaticivorans* DSM 12444 were conducted as previously described in Chapter II. All other chemicals, buffers, and purification reagents were purchased from Sigma-Aldrich, unless otherwise indicated.

Metal Analysis. The metal content of LigW and LigW2 was determined by inductively coupled plasma mass spectrometry (ICP-MS) with an Elan DRC II ICP-MS instrument from Perkin-Elmer. Prior to ICP-MS analysis, loosely bound metals were

removed by passage through a PD-10 desalting column equilibrated with 50 mM HEPES (pH 7.5) previously treated with Chelex 100 ion exchange resin. The protein sample was treated with concentrated HNO_3 for 15 minutes at 100 °C and then diluted with distilled water to a final protein concentration of 1.0 μM and 1% (v/v) HNO_3.
The activation of LigW by Mn^{2+}, Zn^{2+}, Co^{2+}, or Fe^{2+} was analyzed by the addition of these metal ions directly to buffered solutions of the enzyme at pH 7.3 and the catalytic activity was monitored after 12 hours of incubation at 4 °C. Iron was incubated anaerobically under argon prior to measurement of catalytic activity. Metal-free LigW was prepared by dialysis against 10 mM o-phenanthroline at 4 °C in 50 mM HEPES, pH 6.5. A PD-10 column, previously equilibrated with 50 mM HEPES at pH 7.3, was used to remove the o-phenanthroline prior to analysis by ICP-MS.

Site-Directed Mutagenesis. All single site mutations to LigW were constructed following the QuikChange site-directed mutagenesis instruction manual of the manufacturer. The reaction product was transformed into *E. coli* XL1-blue cells and the plasmid isolated using the Wizard Plus SV Miniprep DNA purification system. After verifying the DNA sequence, all mutants underwent standard protein expression and purification as described previously for wild-type LigW (*17*).

Kinetic Parameters. The decarboxylations of 5-carboxyvanillate by LigW and LigW2 were determined by monitoring the decrease in absorbance upon formation of vanillate (VAN) at 312 nm using a differential molar extinction coefficient ($\Delta\varepsilon_{312}$) of 2,040 M^{-1} cm^{-1}. All assays were conducted using a 96-well SpectraMax Plus384 UV-vis spectrophotometer and the steady-state kinetic parameters were obtained using SoftMax Pro 5.0. Standard assay conditions for the decarboxylation of 5-CV included 50 mM

HEPES (pH 7.1), 1.0 mM $MnCl_2$, and 0.1 µM LigW or LigW2 in a final volume of 250 µL at 30 °C. The kinetic constants were obtained using equation 3.1, where v is the initial velocity, E_t is the total enzyme concentration, [A] is the substrate concentration, k_{cat} is the turnover number, and K_m is the Michaelis constant.

$$v/E_t = k_{cat}\,[A] / (K_m + [A]) \qquad (3.1)$$

Measurement of pH-Rate Profiles. The kinetic constants for wild-type LigW and LigW2 k_{cat}, K_m, and k_{cat}/K_m were determined as a function of pH from 5.0 to 9.0. The pH dependence of k_{cat} and k_{cat}/K_m was determined in 50 mM MES (pH 5.5 – 6.75), HEPES (pH 7.0 – 8.25), and CHES (pH 8.5 – 9.0) buffers. The assays were conducted at approximately 0.25 pH unit intervals and the final pH was recorded after completion of the reaction. Differential molar extinction coefficients for the conversion of substrate to product were determined at each pH value; these ranged from ($\Delta\varepsilon_{312}$ = 2,109 M^{-1} cm^{-1}) at pH 5.5 to ($\Delta\varepsilon_{312}$ = 205 M^{-1} cm^{-1}) at pH 9.0. The kinetic data were fit to equation 3.2 using SigmaPlot 11.0, where y is either k_{cat} or k_{cat}/K_m, c is the pH-independent value of y, [H] is the proton concentration, and K_a and K_b are the acid dissociation constants for the ionizing group at low and high pH, respectively.

$$\log y = \log (c /(1+ ([H]/K_a) + (K_b /[H]))) \qquad (3.2)$$

Inhibition Studies. Vanillate (VAN), 3-methoxy-5-carboxybenzoate (MCB), and 5-nitrovanillate (5-NV) were utilized as inhibitors for the reaction catalyzed by LigW. The data for tight-binding inhibition to the enzyme by 5-NV were fit to equation 3.3, where E_t is the total enzyme concentration, [I] is the inhibitor concentration, v_0 is initial enzyme activity in the absence of inhibitor, v_i is the activity of the enzyme in the presence of inhibitor and K_i^{app} is the apparent inhibition constant (*18*). The dissociation

constant, K_d, for the binding of 5-NV to LigW and LigW2 was obtained using equation 3.4, where A is the substrate concentration used in the inhibition experiment and K_m is the Michaelis constant for the substrate 5-carboxyvanillate (5-CV). The inhibition experiments for 5-NV contained 600 μM 5-CV, 50 mM HEPES (pH 7.3), 1.0 mM $MnCl_2$ and 0.1 μM LigW or LigW2 in a total volume of 250 μL at 30 °C. The apparent inhibition constants (K_i^{app}) for VAN and MCB were obtained from a fit of the data to equation 3.5 and the K_d was obtained from equation 3.4. These experiments were conducted at a concentration of 5-CV of 80 μM.

$$v_i/v_\circ = ([E_t] - K_i^{app} - [I] + (([I] + K_i^{app} - [E_t])^2 + (4K_i^{app}[E_t]))^{1/2})/(2[E_t]) \quad (3.3)$$

$$K_i^{app} = K_d\,(1 + [A]/K_m) \quad (3.4)$$

$$v_i/v_\circ = K_i^{app}/([I] + K_i^{app}) \quad (3.5)$$

Solvent Isotope Effects. Steady-state rates for the decarboxylation of 5-CV by LigW were determined in H_2O and D_2O in 50 mM HEPES at pH(D) 7.0, and 1.0 mM $MnCl_2$ at 30 °C. The average of three assays that utilized 0.10 μM LigW in D_2O and H_2O were used to determine the effects on the values of k_{cat} and k_{cat}/K_m. The pH was measured using a Thermo Scientific Orion Star A214 pH/ISE Benchtop Meter with a gel filled pH electrode calibrated with standard buffers. A correction factor of 0.4 was added to meter pH reading in D_2O. (*19*)

Deuterium Exchange Experiments. The enzyme-catalyzed exchange of the hydrogen at C5 of the product VAN with deuterium from solvent was measured as a function of pH. Vanillate (1.9 mM) was incubated in 95% D_2O at pD values of 7.0 (19 mM potassium phosphate) and at 8.0, 9.0, and 10.0 (19 mM NH_4HCO_3) for periods of up to 7 days at 25 ºC after the addition of 48 μM LigW. The reactions were quenched by

the removal of enzyme with a 10 kDa molecular weight cutoff filter and the solvent removed by rotary evaporation. The dried samples were reconstituted with 99% D_2O containing 1.2 mM EDTA to coordinate any paramagnetic metal ions. The fractional exchange of deuterium for hydrogen at C5 of VAN was determined by ^{1}H-NMR spectroscopy. Similarly, the exchange of the hydrogen at C5 of 3-methoxybenzoic acid (1.9 mM) with deuterium from solvent was measured at pD 9.0 in the presence of 48 μM LigW. The apparent first-order rate constants for the exchange of deuterium from solvent at C-5 of VAN were obtained from a fit of the data to equation 3.6, where A_t/A_o is the ratio of VAN that is protonated at time t and k_{obs} is the apparent first-order rate constant for the exchange reaction.

$$\ln[A_t/A_o] = -k_{obs}t \quad (3.6)$$

Product Isotope Effect. The product isotope effect (PIE) was measured in a 50:50 mixture of H_2O and D_2O after the addition of 65 μM LigW to a solution containing 2.0 mM 5-CV at pH/pD values of 7.0 (20 mM potassium phosphate) and 9.0 (20 mM NH_4HCO_3). The reactions were allowed to continue for 1 hour, the enzyme was removed by filtration and then the solvent removed by evaporation. The fraction of hydrogen and deuterium at C5 was determined by integration of the NMR signal for the hydrogen at C5 relative to the nonexchangeable hydrogens at C2 and C6 of VAN.

Membrane Inlet Mass Spectrometry for Detection of CO_2. Membrane inlet mass spectrometry (MIMS) was used to measure the initial rate of CO_2 formation in the reaction catalyzed by LigW (*65*). The method utilizes a permeable membrane inlet linked to an Extrel EXM-200 quadrupole mass spectrometer. The inlet probe was

immersed in a 2.0 mL air-tight reaction vessel containing 50 mM MES (pH 5.5), 1.0 mM $MnCl_2$, and 300 µM 5-CV. The solution was stirred at room temperature for 1.0 minute and the reaction was initiated by the addition of 1.0 µL (183 nM final concentration) of LigW to the reaction vessel. The signal for CO_2 was monitored at an m/z 44 and also at an m/z 28 (as a control). Equivalent experiments were performed with 10 µM 5-CV. The ion current signal was converted to the concentration of CO_2 by means of a standard curve using increments of K_2CO_3 into concentrated acetic acid at room temperature.

Sequence Similarity Network for COG2159. Proteins belonging to COG2159 were identified in the NCBI protein database using the query "COG2159" (*16*). To visualize the phylogenetic relationship amongst proteins in COG2159, sequence similarity enables a comprehensive visualization of related enzymes based on pairwise sequence identity at a given stringent threshold or BLAST *E*-value (*28*).

RESULTS

Purification and Properties of LigW and LigW2. LigW from *Sphingomonas* sp. SYK-6 was purified to homogeneity and the as-purified enzyme exhibited catalytic activity for the decarboxylation of 5-CV to VAN (**Figure 3.2**) with values of k_{cat} = 0.10 ± 0.01 s^{-1} and k_{cat}/K_m = (1.0 ± 0.1) x 10^3 M^{-1} s^{-1}. In contrast, LigW2 from *Novosphingobium aromaticivorans* DSM 12444 performed the reaction with k_{cat} = 18 ± 0.5 s^{-1} and k_{cat}/K_m = 8.4 ± 0.4 x 10^4 M^{-1} s^{-1} and contained 0.2 equivalents of Mn^{2+} in the active site. However, the as-purified LigW contained less than 0.1 equiv. of Mn^{2+} and instead the enzyme retained 0.1 equiv. of each Ni^{2+}, Cu^{2+}, and Zn^{2+} per subunit as determined by inductively coupled plasma mass spectrometry (ICP-MS). The catalytic activity of LigW and LigW2 were abolished after incubation with 10 mM *o*-

phenanthroline at pH 7.5 for 24 hrs. When Mn^{2+} was added directly to the reaction mixture, there was a substantial increase in catalytic activity (**Figure 3.3**). Therefore, LigW and LigW2 were subsequently expressed in *E. coli* in the presence of 1.0 mM $MnCl_2$ that was added directly to the growth medium. The proteins were purified using immobilized Ni ion affinity chromatography and fractions containing LigW and LigW2, respectively were pooled and dialyzed to remove excess imidazole. The metal content of purified LigW that was expressed in the presence of added $MnCl_2$ and then dialyzed against 50 mM HEPES (pH 7.4) contained an average of 0.5 equivalents of Mn^{2+} per monomer with catalytic constants of k_{cat} = 0.9 ± 0.02 s^{-1} and k_{cat}/K_m = (1.0 ± 0.04) x 10^4 M^{-1} s^{-1}.

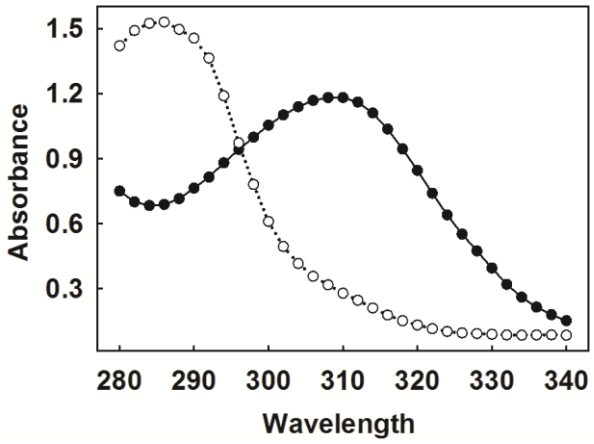


Figure 3.2: UV-Vis Spectrum of 5-carboxyvanillate (—•—) and vanillate (·····○·····).

Dialysis of the Mn^{2+}-substituted LigW against a solution containing 1.0 mM $MnCl_2$, yielded protein with an average of 1.0 equivalents of Mn^{2+} per subunit. The kinetic constants (k_{cat} = 2.2 ± 0.1 s^{-1} and k_{cat}/K_m = 4.0 ± 0.3 x 10^4 M^{-1} s^{-1}) for the decarboxylation of 5-carboxyvanillate by LigW, prepared under these conditions, were obtained using assays supplemented with 1.0 mM $MnCl_2$. Similarly, LigW2 insolated under Mn^{2+} expression conditions and further supplemented with 1 mM $MnCl_2$ in the reaction buffer contained an average of 1.0 equiv. of Mn^{2+}/subunit and decarboxylated 5-CV with k_{cat} = 30 ± 1.3 s^{-1} and k_{cat}/K_m = 3.0 ± 0.3 x 10^5 M^{-1} s^{-1}. The values for LigW

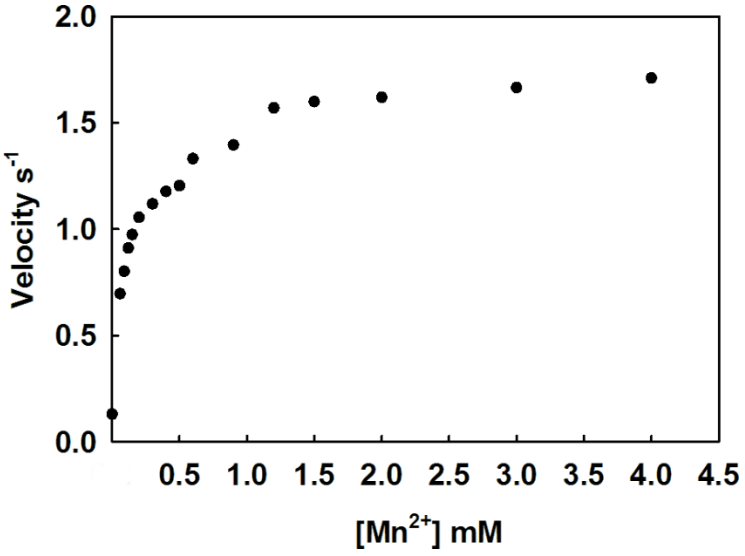


Figure 3.3: Activation of LigW by the addition of Mn^{2+} directly to the assay mixture. The assays were performed using 0.5 μM LigW, 750 μM 5-CV, and 50 mM HEPES at pH 7.4.

are 10-fold greater than for the enzyme that was expressed in the absence of added Mn^{2+} to the growth medium. In contrast, the catalytic constants for LigW2 increased approximately two-fold for k_{cat} and by three-fold for k_{cat}/K_m, when the Mn^{2+}-expression and purification protocols were followed and the protein was furhter supplemented with excess Mn^{2+} in the reaction buffer. In addition, the catalytic activity of LigW2 was inhibited by the presence of Zn^{2+} or Ca^{2+} with K_i of 1.6 ± 0.2 mM and K_i of 47 ± 5.0 mM, respectively. LigW was also deactivated by the presence of excess Zn^{2+}; however, it retained the activity (k_{cat} = 2.8 ± 0.1 s^{-1} and k_{cat}/K_m = 4.0 ± 0.3 x 10^4 M^{-1} s^{-1}) in presence 1 mM $CoCl_2$ in the assay buffer. The catalytic activity for the decarboxylation of MCB by LigW was addressed by monitoring the change in absorbance at 300 nm for the formation of 3-methoxybenzoate. This compound is not a substrate for LigW and the upper limit for k_{cat} is <1.0 x 10^{-3} s^{-1}.

Detection of CO_2. Membrane inlet mass spectrometry (MIMS) was used to differentiate between the formation of CO_2 and HCO_3^- as the initial reaction product in the reaction catalyzed by LigW. The enzymatic reaction was initiated by the addition of LigW to a mixture containing 50 mM MES (pH 5.5), 1.0 mM $MnCl_2$, and 300 μM 5-CV. The time course for the appearance of CO_2, compared to the control reaction without enzyme, is presented in **Figure 3.4**. There was no apparent lag in the formation of CO_2 and the initial velocity for the production of CO_2 was calculated to be 16.3 ± 0.3 μM/min. This rate corresponds to a k_{cat} of approximately 1.5 s^{-1}, a value that is consistent with the value determined spectrophotometrically at this pH.

Inhibition. Vanillate (VAN), 3-methoxy-5-carboxybenzoate (MCB), and 5-nitrovanillate (5-NV) were tested as inhibitors of the reaction catalyzed by LigW (Figure 3.5). The nitro-analogue of the substrate (5-NV) is a very potent inhibitor of LigW with an apparent inhibition constant, K_i^{app}, of 17 ± 7 nM (**Figure 3.5A**). The product VAN and an analogue of the substrate that lacks the phenolic substituent at C4 (MCB) were much poorer inhibitors with apparent K_i values of 157 ± 11 µM and 960 ± 92 µM, respectively (**Figure 3.5B and C**). The dissociation constants for the LigW-inhibitor

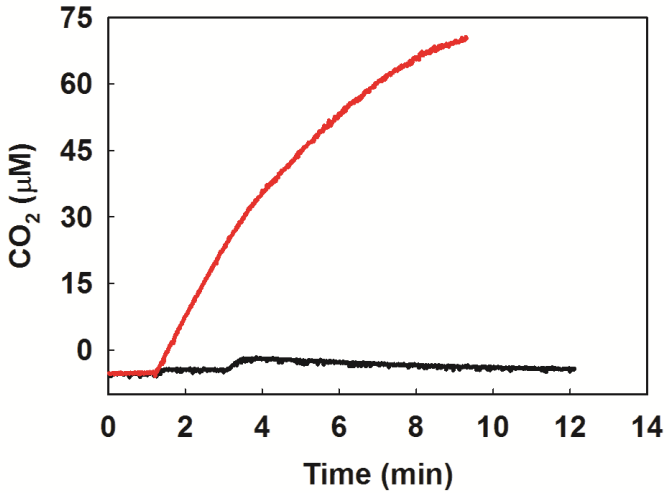


Figure 3.4: Time course for the formation of CO_2 by membrane inlet mass spectrometry. The reaction solution contained 50 mM MES (pH 5.5), 1.0 mM $MnCl_2$, and 300 mM 5-CV. The reaction was initiated by the addition of 183 nM LigW. The red line represents the change in CO_2 concentration after the addition of LigW at 1.0 minute. The black line represents a control sample that did not contain the substrate 5-CV.

complexes, after correcting for the concentration of substrate used in the inhibition experiments, were determined to be 64 ± 6 μM, 390 ± 60 μM, and 1.4 ± 0.7 nM for VAN, MCB, and 5-NV, respectively. The nitro-analogue of the substrate is also a potent inhibitor of the reaction catalyzed by LigW2 with a K_i^{app} value of 17 ± 4 nM (**Figure 3.5D**) and similarly the dissociation constant for binding of 5NV to LigW2 was 2.5 ± 1.0 nM.

The UV-Vis spectrum of the complex formed between LigW and 5-NV was determined by mixing 88 μM 5-NV and 77 μM LigW in a volume of 1.0 mL at pH 7.0. In the absence of LigW the solution of 5-NV is yellow with an absorbance maximum at 430 nm and extinction coefficient of 4102 $M^{-1}cm^{-1}$. Shown in **Figure 3.6A** is the UV-vis spectrum of 5-NV in the presence and absence of LigW. In the presence of LigW the absorbance maximum increased to 470 nm, the extinction coefficient changed to 5213 M^{-1} cm^{-1}, and the solution turned brown in color. The LigW/5-NV complex was heated to 80 ^oC for 15 minutes to denature the protein. The denatured protein was removed by filtration and the spectrum of the filtrate was found to be identical to that of 5-NV in the absence of added LigW. Titrations of LigW from 0 to 88 μM were conducted against a fixed concentration of 88 μM 5-NV at pH 7.0 (**Figure 3.6B**). No absorbance change occurred when the titration was initiated with LigW without Mn^{2+} in the active site. Likewise there was no observable change in absorbance, when 5-NV was incubated with Mn^{2+} in absence of LigW.

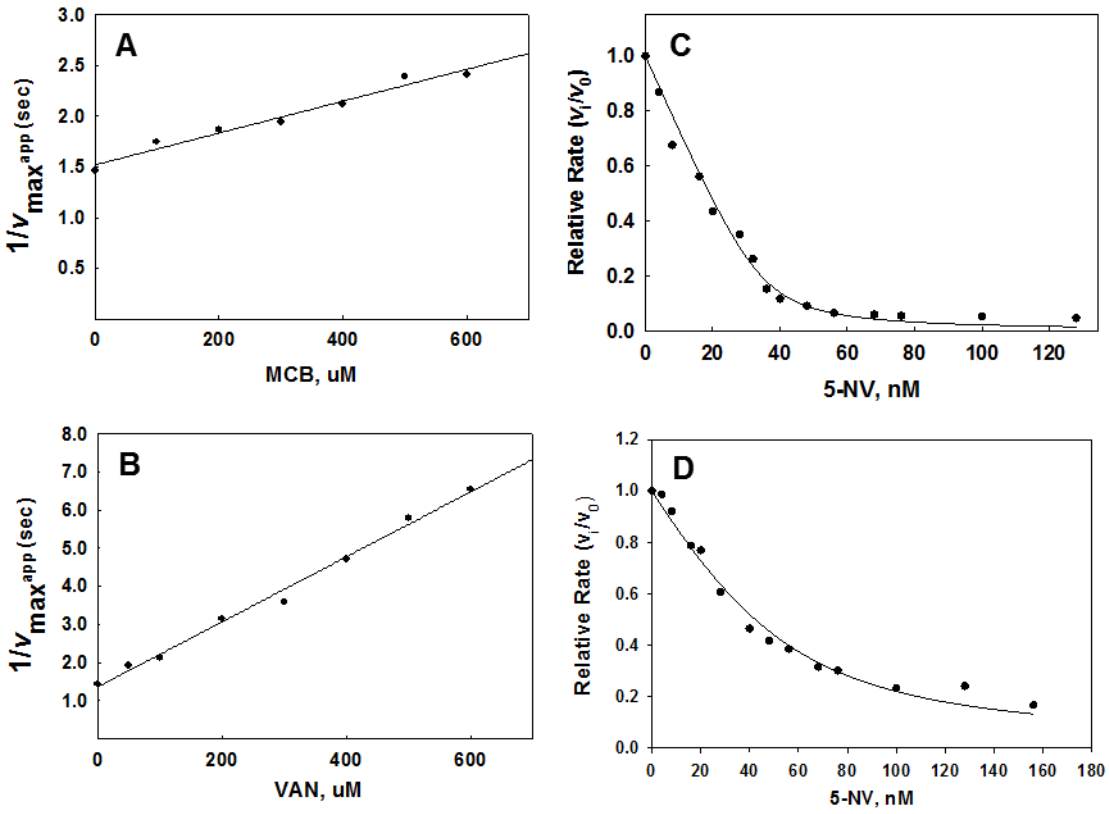


Figure 3.5: Inhibition of the reaction catalyzed by LigW by MCB (A) and VAN (B) with data fitted to equation 3.5. LigW (C) and LigW2 (D) inhibition by 5-NV. The data were fit to equation 3.3.

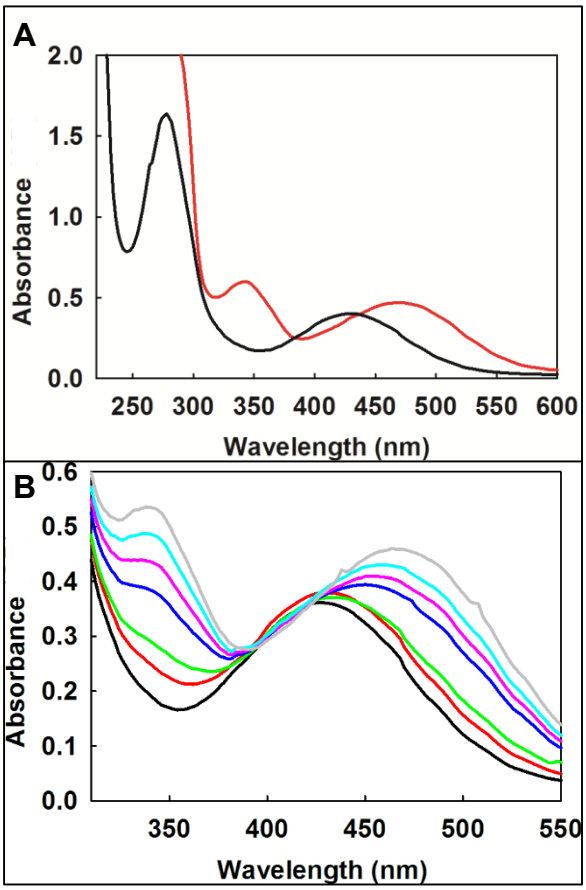


Figure 3.6: All reactions were recorded in 1 cm path length cuvette containing 50 mM HEPES (pH 7.0) and 1 mM $MnCl_2$. Recorded spectra in the range of 220 - 600 nm of the following: red (88 µM 5-NV + 77 µM Mn^{2+} • LigW), black (88 µM VAN), and black (88 µM 5-NV) (**A**). UV-Vis spectra of various concentrations 0 (black), 10 (red), 20 (green), 40 (blue), 50 (purple), 60 (cyan), and 70 (gray) µM of Mn^{2+} • LigW (**B**).

Mutation of Active Site Residues. Ten residues localized in the active site of LigW (Glu-7, Ala-9, Arg-45, His-173, Tyr-197, His-226, Glu-229, Arg-237, Asp-296, Tyr-299) were mutated and the purified proteins tested for changes in catalytic activity (**Figure 3.7**). Modification of any of the three residues that bind the divalent cation in the active site (Glu-7, His-173, or Asp-296) diminished the value of k_{cat}/K_m by at least three orders of magnitude, relative to that of the wild-type enzyme. His-226 and Glu-229 form an apparent hydrogen bond network to the invariant aspartate (Asp-296) at the end of β-strand 8. Mutation of these residues to asparagine and glutamine, respectively, lowers the values of k_{cat} and k_{cat}/K_m by approximately three orders of magnitude. The mutation of Ala-9 to histidine resulted in a relatively modest reduction in the catalytic activity of LigW. In most members of the AHS this residue position is occupied by a histidine residue, which coordinates the active site metal ion. In the structure of LigW Wat-3 is ligated to the metal ion at this location.

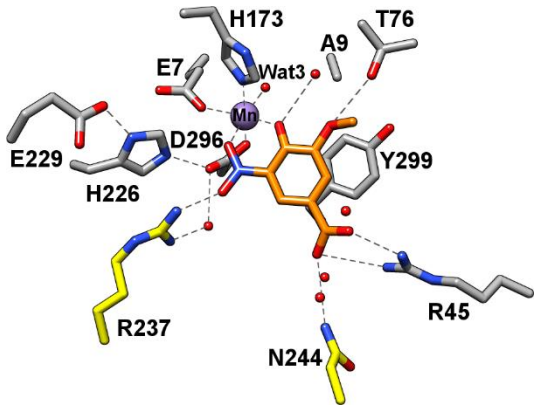


Figure 3.7: Active site of LigW in the presence of the inhibitor 5-nitrovanillate (taken from PDB id: 4NG3).

Mutation of the tyrosine residues (Tyr-197 and Tyr-299) that interact with the aromatic core of the substrate diminished the activity of LigW significantly. The two arginine residues (Arg-237 and Arg-47) that ion-pair with the two carboxylate groups of the substrate cannot be mutated to alanine without the loss of catalytic activity. The kinetic constants for the decarboxylation of 5-CV by the LigW mutants are presented in **Table 3.1**. In addition, the kinetic constants for LigW2 D314N mutant were determined to be as follows: k_{cat} = 0.017 ± 0.002 s^{-1}, K_m = 212 ± 57 μM, and k_{cat}/K_m = 79 ± 13 M^{-1} s^{-1}, indicating that the D314N ligW2 mutation is disruptive for the catalytic step as the value of k_{cat} drops over 3 orders of magnitude for the mutant enzyme as compared to the wild-type protein. However, the K_m value is not affected by this mutant, when compared to the wild-type enzyme. Aspartate 314 to asparagine mutant was chosen, because this residue is highly conserved in all functionally characterized AHS members and it is known to activate the hydrolytic water for catalysis in those reactions (*6*).

Table 3.1: Kinetic constants for LigW and various mutants at pH 7.0 and 30 ^{o}C.

Enzyme	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1} s^{-1}$)
wild-type	2.24 ± 0.03	55 ± 4	(4.0 ± 0.6) x 10^4
E7Q	(8.0 ± 0.8) x 10^{-4}	33 ± 12	24 ± 7
E7A	(4.6 ± 1.5) x 10^{-3}	430 ± 260	11 ± 3
A9H	1.2 ± 0.1	32 ± 5	(3.7 ± 0.8) x 10^4
R45A	(2.7 ± 0.2) x 10^{-2}	700 ± 90	38 ± 2
H173N	(7.3 ± 0.2) x 10^{-2}	760 ± 30	96 ± 2
Y197A	(2.1 ± 1) x 10^{-3}	1200 ± 700	2.0 ± 0.3
H226N	(1.4 ± 0.2) x 10^{-3}	190 ± 60	7.6 ± 1.6
E229Q	(5.0 ± 0.2) x 10^{-3}	58 ± 9	87 ± 11
R237A	(7.2 ± 2.9) x 10^{-3}	612 ± 420	12 ± 4
D296N	(3.2 ± 0.4) x 10^{-3}	460 ± 0.90	7 ± 1
Y299A	(8.0 ± 2.0) x 10^{-4}	130 ± 0.80	6.4 ± 2.6

pH-Rate Profiles. The kinetic constants for the decarboxylation of 5-CV by LigW were determined as a function of pH (**Figure 3.8**). For k_{cat}, the pH-rate profile is bell-shaped with pK_a and pK_b values of 4.3 ± 0.1 and 8.3 ± 0.1, respectively. Similarly, for the k_{cat}/K_m profile the catalytic activity is diminished at high and low pH. The kinetic pK_a and pK_b values determined for the effects on k_{cat}/K_m are 4.7 ± 0.2 and 7.6 ± 0.1, respectively.

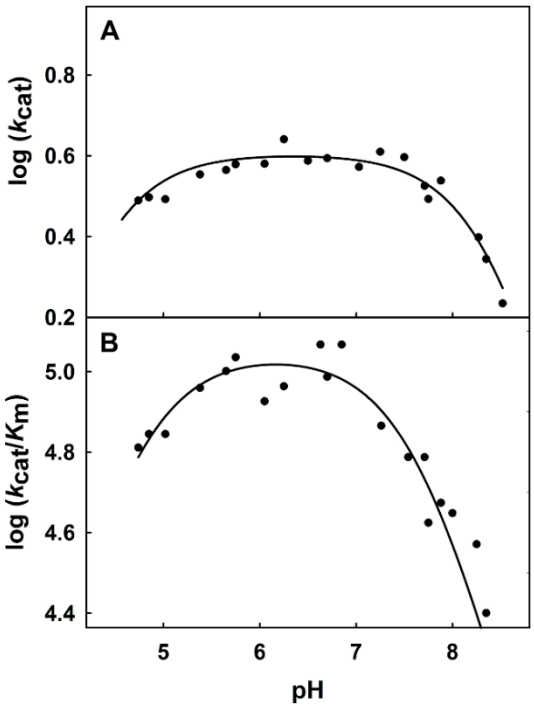
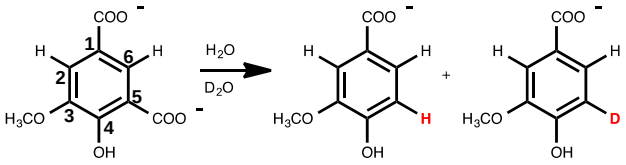


Figure 3.8: The pH-rate profiles for the decarboxylation of 5-CV by LigW. The data for the changes in k_{cat} (panel **A**) and k_{cat}/K_m (panel **B**) were fit to equation 3.2.

Solvent Isotope Effects. The solvent isotope effects were determined at pH(D) 7.0 by direct comparison of the rates in H_2O and D_2O. The measured solvent isotope effects on k_{cat} and k_{cat}/K_m at pH(D) 7.0 are 2.9 and 2.0, respectively. The product isotope effect (PIE) was determined by measuring the ratio of deuterium to hydrogen in the product VAN, when the decarboxylation of 5-CV by LigW was conducted in an equal mixture of H_2O and D_2O at a pD/pH of 9.0 (**Scheme 3.1**). The ratio of H:D at C5 was determined from the integrated signal intensity of the hydrogen at C5 relative to the non-exchangeable hydrogens at C2 and C6 (**Figure 3.9**). The 50:50 H_2O and D_2O reaction mixture was allowed to reach completion in approximately 1 hour and the ^{1}H-NMR spectrum was compared to the spectrum of a control reaction conducted in H_2O (**Figure 3.10A and B**). The integrated area (0.82 units) of the C-5 proton represents the remaining 5-^{1}H VAN after the reaction is complete. Thus, the PIE was calculated to be 4.6 with an estimated error of ± 0.01, when compared to the integrated area of the resonances at the C2 and C6 positions.



Scheme 3.1: Schematic representation reaction of VAN in presence of ligW in 50:50 H_2O:D_2O.

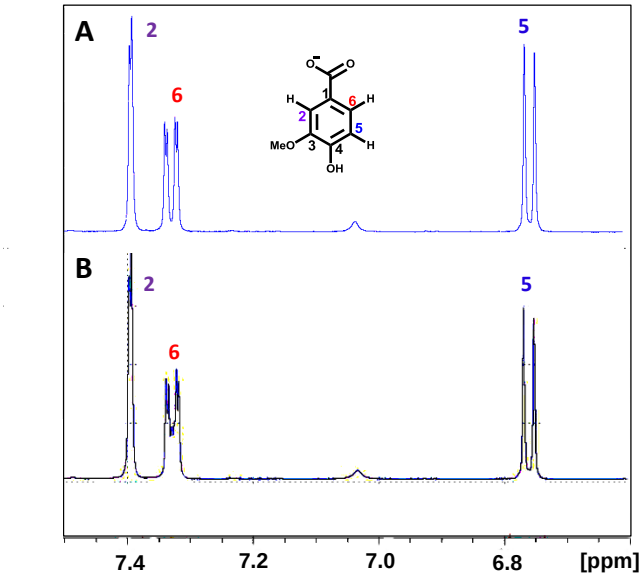
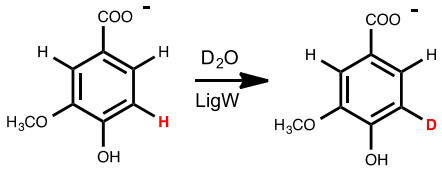


Figure 3.9: ^{1}H-NMR spectrum of the control reaction carried out in H_2O with 2 mM 5-CV, 65 µM LigW in in NH_4HCO_3 buffer at pD 9.0 allowed to react for 1 hour (**A**). ^{1}H-NMR spectrum of 2 mM 5-CV, 65 µM LigW in in NH_4HCO_3 buffer at pD 9.0 allowed to react for 1 hour (**B**). The integrated area of the C-5 proton signal is used to calculate PIE.

Deuterium Exchange Experiments. The enzyme-catalyzed exchange of deuterium from solvent with the hydrogen at C5 of the product VAN was determined by ^{1}H-NMR spectroscopy (**Scheme 3.2**). The initial reaction mixture contained 2.0 mM VAN and 65 μM LigW in D_2O at pD values of 7.0, 8.0, 9.0, and 10.0. The solvent exchange reaction at C5 of VAN resulted in the diminution of the NMR signal for the C5 hydrogen at 6.78 ppm and a loss of coupling to the hydrogen at C6 at 7.33 ppm (**Figure 3.10A and B**). The observed first order rate constants (k_{obs}) for the solvent exchange reactions at pD 7.0, 8.0, 9.0, and 10.0 were (5.3 ± 0.9) x 10^{-5} s^{-1}, (1.0 ± 0.7) x 10^{-4} s^{-1}, (2.0 ± 0.1) x 10^{-4} s^{-1}, and (2.0 ± 0.3) x 10^{-5} s^{-1}, respectively. These values were calculated by plotting the natural log of the ratio of the concentration of 5-^{1}H VAN at time (t) and the initial concentration of 5-^{1}H VAN versus time in hours. The slope of the line represents the observed first order rate constants (k_{obs}) shown in **Figure 3.11**. The rate constants for the exchange reaction (k_{ex}) that corrected for the concentration of enzyme and product used in the experiment were calculated using equation 3.7. The corrected values of k_{ex} are 2.1 x 10^{-3}, 3.9 x 10^{-3}, 7.9 x 10^{-3} and 7.9 x 10^{-4} s^{-1} at pH values of 7, 8, 9, and 10, respectively. No exchange (<10%) was catalyzed by LigW with 3-methoxybenzoate after an incubation period of 7 days. The catalytic activity of LigW



Scheme 3.2: Representation of the deuterium exchange reaction at C-5 of VAN.

was not diminished for up to seven days. In addition, a reaction in a 50:50 mixture of

D_2O:H_2O was further performed with 2 mM VAN, 65 µM LigW, and 20 mM phosphate buffer at pD 7.0 for nine days. The resulting ^{1}H-NMR spectrum revealed the integrated area of C-5 proton to be 0.52 units, suggesting that 50 % of the VAN was in the 5-^{2}H VAN form (**Figure 3.12 A and B**).

$$k_{ex} = k_{obs}([VAN]/[LigW]) \quad (3.7)$$

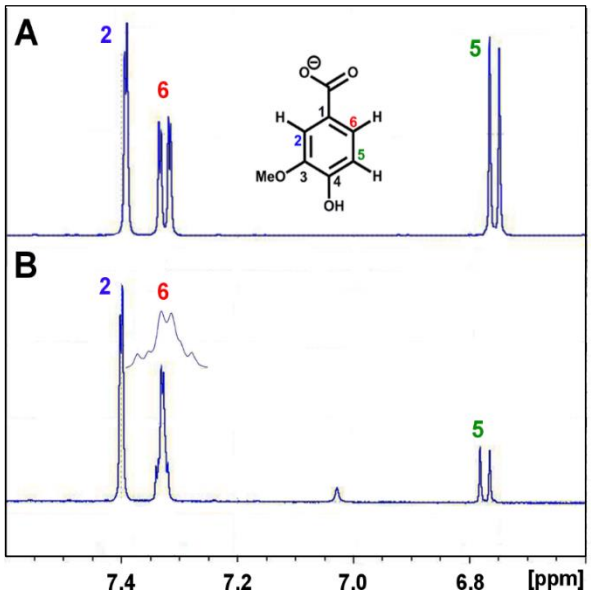


Figure 3.10: Enzymatic deuterium exchange at C-5 of VAN examined by ^{1}H-NMR spectroscopy. ^{1}H-NMR spectrum of the control in D_2O of 2 mM VAN in NH_4HCO_3 buffer at pD 9.0 in absence of LigW for nine days. The benzene ring of VAN molecule is numbered for clarity (**A**). The reaction mixture in D_2O of 2 mM VAN and 65 µM LigW in NH_4HCO_3 buffer at pD 9.0 for nine days. Proton signals at C-2, C-6, and C-5 are shown with a close up of the C-6 multiplet resulting from the deuterium exchange at C-5 (**B**).

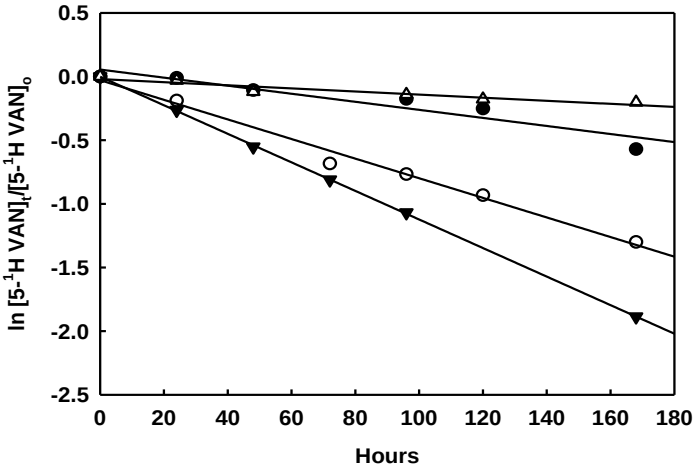


Figure 3.11: Deuterium exchange of the C-5 proton in VAN in presence of LigW at pD 7.0, 8.0, 9.0, and 10.0 over time. The linear regression of ln ($[5\text{-}^{1}H\ VAN]_t/[5\text{-}^{1}H\ VAN]_0$) over time in hours yields a slop of a line with a - *k* or the first order rate constant at the given pD value. The data were fitted to equation 3.6.

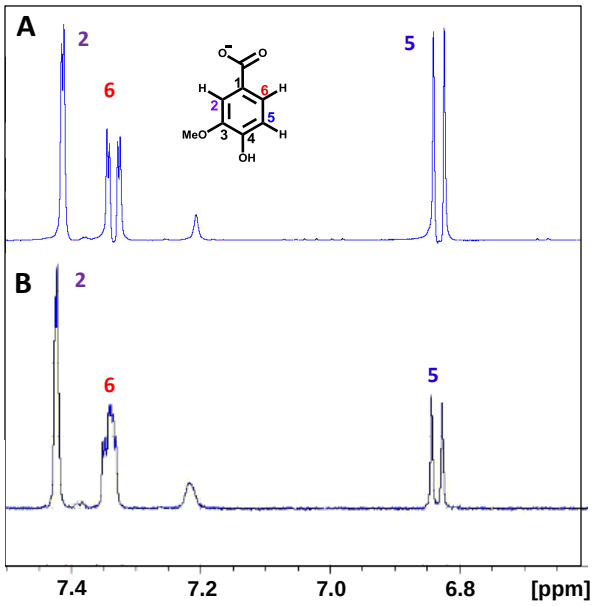
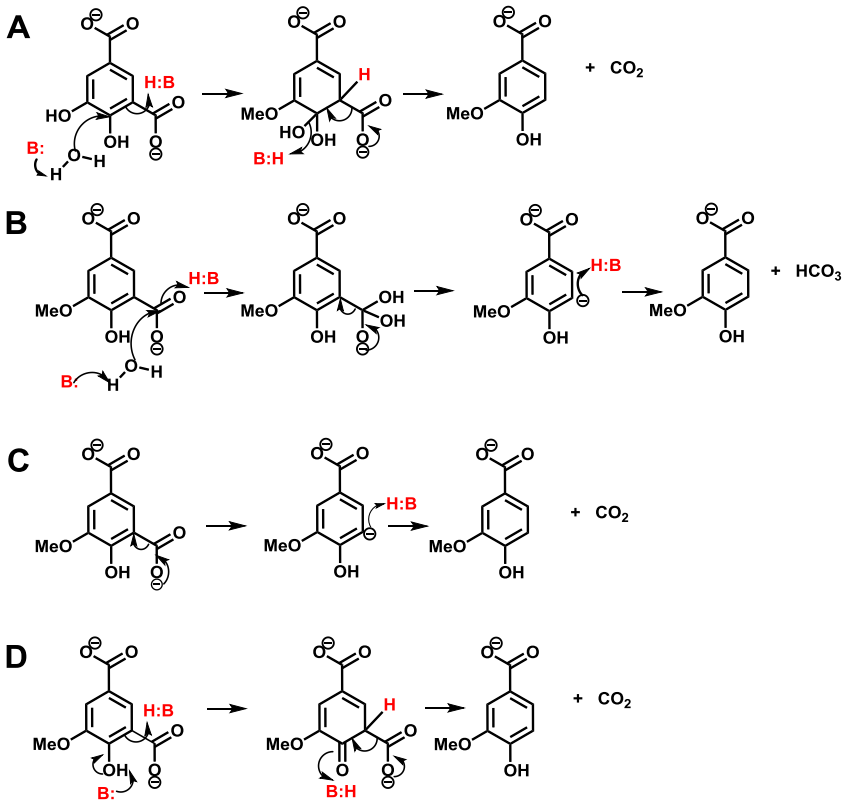


Figure 3.12: ^{1}H-NMR spectrum of 2 mM VAN in H_2O control (**A**). ^{1}H-NMR spectrum of 50:50 D_2O:H_2O mixture of 20 mM phosphate buffer (pD 7.0) and 65 µM LigW allowed to stand in room temperature for nine days (**B**).

DISCUSSION

Decarboxylation Reaction Mechanisms. Lig W can catalyze the decarboxylation of 5-carboxyvanillate using a range of potential reaction mechanisms, and examples of these transformations are illustrated in **Scheme 3.3**. The decarboxylation of 5-carboxyvanillate requires the cleavage of the carbon-carboxylate bond at C5 and formation of a new carbon-hydrogen bond at the same carbon. The primary issues are whether C-H bond formation precedes or follows the cleavage of the carbon-carboxylate bond, and whether the initial reaction product is carbon dioxide or bicarbonate. Since LigW is a member of the amidohydrolase superfamily, this enzyme



Scheme 3.3: Possible reactions for the decarboxylation of 5-CV by LigW.

is homologous to a substantial number of enzymes that catalyze the hydrolysis of carboxylate esters and amides via the activation of a water molecule via coordination to a divalent cation within the active site. The high-resolution crystal structure of LigW in the presence and absence of bound inhibitors has clearly demonstrated that this enzyme contains a divalent metal ion in the active site and the experiments reported in this chapter demonstrate that manganese is required for catalytic activity. In the absence of a bound ligand the metal ion is coordinated to three water molecules and the side chains of Glu-7, His-173, and Asp-296. When the tight-binding inhibitor 5-NV binds to the enzyme two of the water molecules are displaced (Wat-1 and Wat-2) and the metal ion coordinates the phenolic group from C4 and the nitro group at C5.

In mechanism A, the reaction is initiated by the hydration of the carbon-carbon double bond between C4 and C5. The hydrated intermediate subsequently collapses via the cleavage of the carbon-carboxylate bond at C5 and the newly formed carbon-hydroxyl bond at C4 to generate the product vanillate. In mechanism B, an activated water molecule attacks the carboxylate carbon to form a bicarbonate-like intermediate, which then cleaves to form bicarbonate and an unstable carbanion at C5, which is subsequently protonated by an active site acid. It is has been reported that acid catalyzed decarboxylation reactions of aromatic compounds are more energetically favorable via addition of a water molecule to the leaving carboxylate and formation of protonated carbonic acid prior to cleavage of the C-C bond (*66*). If an activated water molecule is not employed in the reaction mechanism catalyzed by LigW, then an alternative mechanism can be proposed (mechanism C), where CO_2 is the initial reaction product and the unstable carbanion at C5 is protonated by an active site general acid.

Alternatively, the lone pair of electrons from the ionized phenolic group can delocalize to the carbon at C5 and be protonated by an acid in the active site. Decarboxylation of this intermediate would lead directly to the vanillate product (mechanism D).

The three-dimensional structure of LigW, determined with the tight-binding inhibitor 5-nitrovanillate (5-NV) in the active site, diminishes the likelihood that an activated water molecule is involved in the decarboxylation of 5-carboxyvanillate. When 5-NV binds to the active site of LigW, two of the three water molecules (Wat-1 and Wat-2) that are initially bound to the manganese in the active site are displaced. The remaining water molecule (Wat-3) is 4.6 Å from C4 and 5.1 Å from the nitro-group of the inhibitor is ill positioned to attack either the aromatic ring itself or the C5-carboxylate group of the substrate (**Figure 3.7**). Moreover, the high-resolution crystal structure of LigW does not reveal any other nucleophiles that are positioned to add to either C4 or the C5-carboxylate of the substrate. The MIMS experiment is fully consistent with the initial release of CO_2 from the active site and thus the carboxylate group is unlikely to be attacked by an activated water molecule or a protein-based nucleophile. Therefore, neither mechanism A nor mechanism B, depicted in **Scheme 3.3**, is likely to be catalyzed by LigW. These considerations support a mechanism where protonation at C5 either precedes or follows decarboxylation of the substrate.

Protonation of the C5 Carbon. The four X-ray structures of LigW that are reported in the previous chapter support the proposal that the hydrogen at C5 of the product vanillate is donated either from Asp-296 or the phenolic group at C4 of the substrate. In the structure of LigW with 5-NV, one of the two carboxylate oxygens from the side chain of Asp-296 is 4.0 Å away from C5 of the inhibitor and in the

corresponding structure with MCB in the active site, this same oxygen atom is 4.3 Å from C5 and in the structure with the product vanillate, it is 3.7 Å. The closest other entity that could function as the proton donor is the phenolic substituent from C4. Asp-296 is conserved in all of the enzymes that have been functionally characterized in the amidohydrolase superfamily, and in those enzymes that catalyze hydrolytic reactions this residue has been demonstrated to function as a general acid during the protonation of the leaving group (*7*). Mutation of this residue to an asparagine in LigW results in the reduction in catalytic activity by more than three orders of magnitude and thus this residue is critical for the catalytic activity of LigW.

The most remarkable finding from the structures of LigW presented in the previous chapter is the distortion of the two substituents attached to C4 and C5 of the bound ligand. The out-of-plane bending of the carboxylate substituent at C5 would facilitate the protonation of this carbon from the *si*-face of C5 from either the hydroxyl group at C4 or the carboxylate group of Asp-296. The pK_a of the phenolic group of the substrate is 11.7 but the complexation of the substrate to the divalent cation in the active site will significantly reduce the pK_a of the phenolic group of the bound substrate.

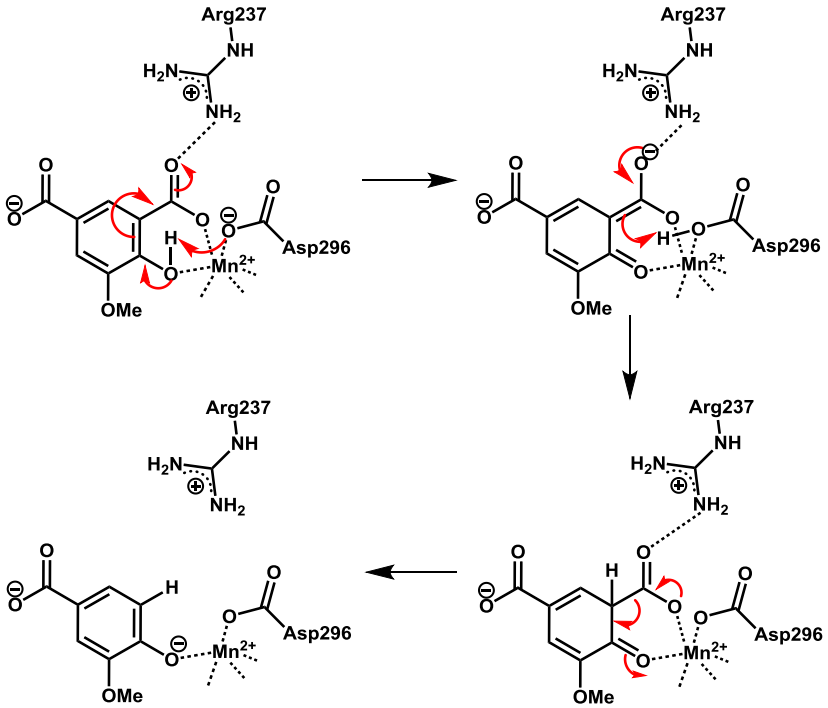
Timing of Proton Transfer. In reaction mechanisms C and D, the proton that is ultimately found at C5 of the product vanillate is transferred either after the carboxylate group has been cleaved or before. If protonation occurs after decarboxylation, it is anticipated that this step would be fast relative to the actual decarboxylation step and thus the product isotope effect (PIE), determined using a 50:50 mixture of H_2O:D_2O, should be near unity. Alternatively, if protonation preceded decarboxylation, then it is expected that the protonation step would be more rate-limiting than the decarboxylation

step, and thus the PIE would be measurably greater than unity. With LigW the PIE was determined to be 4.6 and thus we conclude that LigW catalyzes the decarboxylation via a reaction mechanism that resembles mechanism D of **Scheme 3.3**. In the decarboxylation reaction catalyzed by orotidine 5'-monophosphate decarboxylase (OMPDC) the PIE was determined to be 1.0 and the reaction mechanism has been proposed to occur via the formation of an unstable vinyl carbanion intermediate (*67*). In LigW, the critical participation of the phenolic group at C4 is further supported by the fact that we were unable to demonstrate that the substrate analogue that is missing the phenolic substituent (MCB) can be decarboxylated by LigW. The x-ray structure of LigW in the presence of MCB clearly demonstrates that this substrate analogue can bind to the active site of LigW in a manner that is nearly identical to that of the tight binding inhibitor 5-NV and the product vanillate.

We have also demonstrated that LigW will catalyze the exchange of the proton at C5 of the product vanillate with a rate constant of 7.9×10^{-3} s^{-1} at pD 9.0. This value is more than two orders of magnitude greater than the exchange of the proton at C6 of UMP catalyzed by OMPDC at pD 7.58 (*68*). We were unable to observe any exchange with solvent over an incubation time of 7 days using 3-methoxybenzoate as a probe substrate. This observation further supports the proposal that the exchange reaction requires the participation of the phenolic group at C4 via the delocation of the lone pair of electrons from the phenolate to C5.

Mechanism of LigW Catalyzed Reaction. The proposed mechanism of action for the decarboxylation of 5-carboxyvanillate by LigW is presented in **Scheme 3.4**. In this mechanism the binding of the substrate to the active site of LigW results in the

displacement of two water molecules from the mononuclear metal center in the active site of the enzyme. The phenolic group at C4 and the carboxylate group at C5 are directly coordinated to the divalent cation. The formation of this complex is fully supported by the X-ray structures of complexes of LigW crystallized in the presence of 5-nitrovanillate, vanillate, and MCB. The binding of the substrate in the active site is accompanied by a significant out-of-plane bending of the substituents at C4 and C5 together with "splitting" of the benzene ring (see Chapter II) of the substrate and a



Scheme 3.4: Proposed mechanism for the decarboxylation of 5-CV to VAN by LigW.

reduction in the pK_a for the ionization of the phenolic group at C4. The proton from the phenol is subsequently transferred to the side chain carboxylate from Asp-296 or directly to C5 after delocalization of the lone pair of electrons from the ionized phenolate. The proton transfer step is likely to be the rate limiting for the overall reaction, since there is a significant solvent isotope effect on the value of k_{cat} and k_{cat}/K_m, in addition to a substantial product isotope effect, when the reaction is conducted in a mixture of H_2O and D_2O. Decarboxylation follows the protonation of C5, but it is uncertain whether or not the phenolate is protonated before, or after, the product leaves the active site.

Most of the residues contained within the active site of LigW are critical for catalytic activity of the enzyme. Mutation of either of the two arginine residues that ion-pair with the carboxylates at C1 and C5 diminishes the value of k_{cat}/K_m by approximately three orders of magnitude. Mutations of the tyrosine residues that sandwich the aromatic ring of the substrate results in the loss of activity of nearly four orders of magnitude. Mutation of the putative proton transfer residue, Asp-296, diminishes the catalytic activity by nearly four orders of magnitude. The x-ray structure of LigW has revealed a hydrogen bond network from Asp-296 to His-226 and Glu-229. These two residues are also conserved in other decarboxylases from COG2159 and mutation of these residues diminishes the value of k_{cat}/K_m significantly. The specific catalytic function of this hydrogen bonded network is unclear. These residues may modulate the pK_a of Asp-296 or this hydrogen bonded network may function as a conduit for protons to and from the active site. The pH rate profile of the enzyme is bell shaped, where one residue must be ionized for optimal catalytic activity and another residue protonated. The loss of activity at low pH likely reflects the protonation of Asp-296 and the ionization at high pH likely

reflects the ionization of His-226, since this is the only other residue within the active site that is likely to be protonated in this range of pH values.

Implications for Other Decarboxylases from COG2159. The proposed LigW mechanism differs from suggested chemical mechanisms of γ-resorcylate decarboxylase (γ-RSD) (*24*), α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) (*26*), and uracil-5-carboxylate decarboxylase (IDC) (*22*). Based on the available crystal structures of γ-RSD complexed with substrate γ-resorcylate (PDB id: 2DVU) and inhibitor 2, 3-dihydroxybenzaldehyde (PDB id: 2DVX), these ligands bind to the metal center in a monodentate fashion via the C2 hydroxyl group. Thus, the leaving carboxylate group at C1 of the substrate interacts with the arginine residue (Arg-229*) from the adjacent subunit of the γ-RSD homotetramer in a similar fashion as Arg-237* in LigW. The invariant β-strand 8 Asp-287, then abstracts the proton from the metal ligated hydroxyl group with subsequent delocalization of the lone pair at the C1 position of the phenyl ring. The carbanion then abstracts a proton from a nearby water (Wat1) molecule forming a sp^3 hybridized intermediate. Decarboxylation at C1 occurs with concomitant donation of the aspartate proton to the C2 phenolate, resulting in the formation of CO_2 and resorcinol (*24*). Although this mechanism is plausible, it is still unclear whether the crystalline enzyme is active, since it is unable to decarboxylate γ-resorcylate. Furthermore, the water presumably shuttling a proton to the C1 position (3.8 Å) interacts with the arginine from the adjacent subunit and His-218 from β-strand 6, thus lowering its pK_a value and enabling the transfer of one of the protons to the C1 of γ-resorcylate. The fate of the formed hydroxide ion has not been addressed.

To further investigate the substrate specificity and mechanisms of decarboxylases in COG2159, the human ACMSD crystal structure in presence of the inhibitor 1, 3-dihydroxyacetonephosphate (DHAP) (PDB id: 2WM1) and pyridine-2, 6-dicrboxylic acid (PDC) (PDB id: 4IH3) have been compared (*25, 27*). Superposition of the two structures showed that ACMSD is a monomer with DHAP and three pairs of dimers per asymmetric unit with PDC. The active site metal center is ligated by three histidine residues and aspartate from β-strand 8 (*69*). Several mechanisms of the zinc dependent ACMSD have been outlined based on the available crystal structures (*25, 27, 70, 71*). In the ACMSD structure with DHAP, the inhibitor is not coordinated to the metal center and the binding site consists of an arginine residue from the insertion domain (Arg-47), which establishes a salt bridge with the phosphate moiety and hydrogen bonds to the keto group. Additionally, the arginine residue from the adjacent monomer (Arg-235*) in ACMSD is conserved as determined by sequence alignments with COG2159 proteins and it is seen in the PDC bound structure (*25*). The equivalent residues in the structure of ACMSD from *Pseudomonas fluorescens* (PDB id: 2HBV) are Arg-51 and Arg-239* (*71*). Unfortunately, the structure of this enzyme has not been determined in the presence of an inhibitor or other ligand that more resembles the substrate or the decarboxylated product.

The most recent mechanism of ACMSD suggests that the substrate, α-amino-β-carboxymuconate-ε-semialdehyde (ACMS), orients into the active site with the C2 carboxyl group towards the Zn^{2+} ion, so that the leaving carboxylate group at C3 is stabilized by Arg-51 from the insertion domain and away from the metal center. The C2 carboxylate is further interacting with Arg-239* from the adjacent monomer. The zinc

ligated water molecule (Wat1) is then activated by the histidine residue from β-strand 6. A nucleophilic attack by Wat1 is achieved at the C2 position of the substrate with subsequent protonation at C3 by the protonated form of histidine from β-strand 6. Decarboxylation of the C3 carboxyl group is initiated with collapse of the tetrahedral intermediate and release of CO_2 and 2-aminomuconate semialdehyde (AMS) (*71*).

When the mechanisms of ACMSD and IDC are compared, it is seen that the zinc-dependent isoorotate decarboxylase (IDC) presumably utilizes a non-metal bound water for the decarboxylation reaction. The high-resolution crystal structure of IDC was determined previously in the presence of 5-nitro-uracil (PDB id: 4HK6). In IDC, the active site zinc is coordinated to the nitro group and the C4 oxygen of the bound inhibitor. In addition, Arg-262* from the adjacent subunit interacts with the nitro-substituent of the inhibitor in a similar manner to that of Arg-235* in ACMSD and Arg-237* in LigW. The decarboxylation reaction presumably could be accomplished via two proposed routes. In the first mechanism, the substrate presumably binds into the active site replacing the two metal bound waters with the C4 hydroxyl oxygen and C5 carboxylate oxygen in a bidentate fashion. In addition, the highly conserved histidine residues from β-strands 1 and 5 together with aspartate from β-strand 8 are ligated to the Zn^{2+} ion establishing an octahedral geometry. The O-2 and the N-3 form hydrogen bonds with Asn-98. Whereas, the N-1 interacts with Arg-68, which extends from the insertion domain. One of the oxygen atoms of Asp-323 is activated by the metal center, while the second oxygen is hydrogen bonded to His-251 from β-strand 6. A nucleophilic attack by Asp-323 is performed on the carbon center of the C5 carboxyl group. The unstable tetrahedral intermediate collapses concomitantly with protonation at C-5 from

solution forming uracil and a carboxylated aspartate intermediate. The Arg-262* and His251 coordinated hydroxide could nucleophilically attack at the leaving carboxyl group or the side chain carboxylate of the Asp-232 intermediate, releasing HCO_3^-. The second mechanism, involves the addition of the hydroxide to the carbon center of the leaving carboxylate, forming an unstable tetrahedral intermediate, which is stabilized by Arg-262* and the metal center. Collapse of this intermediate leads to the formation of the product uracil and HCO_3^- (*22*).

Structurally and mechanistically characterized COG2159 decarboxylases have unveiled the minimal catalytic form of these enzymes is a homodimer. The arginine residue perturbing into the active site from the adjacent monomer is highly conserved across all the decarboxylases. In addition, amino acid residues contained in the insertion domains between β - strands 1 and 2 of the $(\beta/\alpha)_8$ -TIM barrel fold have been established to be important for substrate binding and orientation and vary among the decarboxylases. Based on the crystal structure of γ-RSD with the substrate γ-resorcylate, it is likely that the crystalline enzyme is not active. Evidence of a bidentate ligation to the metal center is seen in the 1.6 Å high resolution crystal structure of γ-RSD with a nitro analogue, 2-nitroresorcinol (PDB id: 4QRO). This inhibitor has apparent Ki value of 96 ± 8.1 μM.

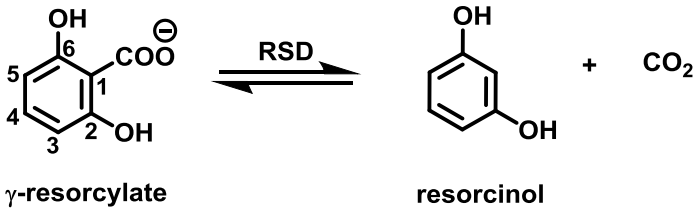
Furthermore, the recently published mechanism of ACMSD suggests that the leaving carboxylate group of ACMS is positioned toward Arg-47, which is inconsistent with the proposed mechanisms for γRSD, IDC, and LigW. The nitro analogue of the IDC substrate further supports that the 5-carboxyl-uracil binds through the leaving carboxylate and adjacent oxygen atoms to the metal center. However, the mechanism of IDC involves either a covalent intermediate with the aspartate from β-strand 8 or a direct

hydrolytic attack of a hydroxide on the carbon center of the leaving carboxylate. Either of these mechanisms result in the formation of bicarbonate as an initial product. Evidence of a covalent intermediate with the metal bound aspartate have not been reported for enzymes in the amidohydrolase superfamily (*7*). The AHS relies predominately on metal-assisted hydrolytic reactions; thus, Wat 3 is unlikely to add to the C5 carboxylate. Further studies are necessary to establish the mechanisms of COG2159 decarboxylases; however, the elucidation of the chemical mechanisms of LigW could shed some light on the possibility of these decarboxylation reactions.

CHAPTER IV

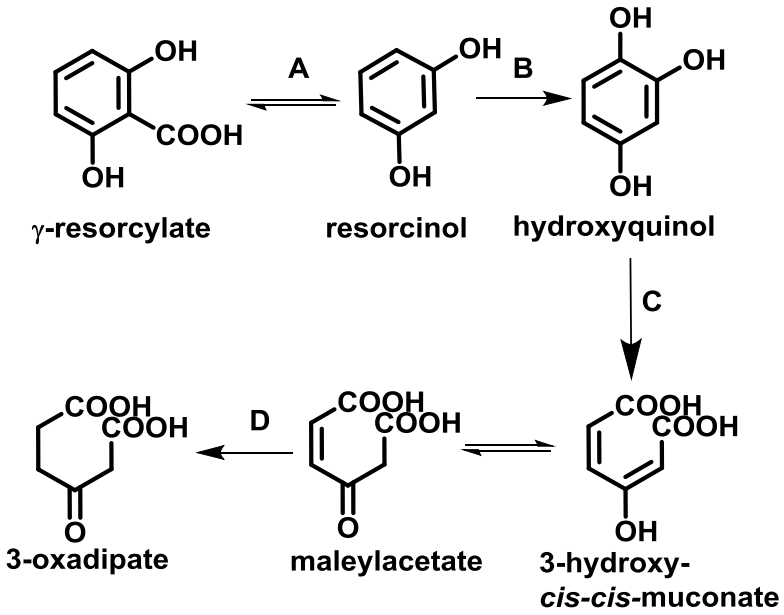
γ-RESORCYLATE DECARBOXYLASE (γRSD) AND OTHER COG2159 ENZYMES

γ-Resorcylate decarboxylase (γRSD) from *Rhizobium* sp. MTP-10005 is the closest structurally characterized homologue of 5-carboxyvanillate decarboxylases (LigW) from *Sphingomonas* sp. SYK-6 and LigW2 from *Novosphingobium aromaticivorans* DSM 12444 sharing 33% sequence identity with both enzymes. This decarboxylase catalyzes the C-C bond cleavage at C1 of γ-resorcylate (γ-RS; 2, 6-dihydroxybenzoic acid) to resorcinol (1, 3-dihydroxybenzene) as shown in **Scheme 4.1** (*72, 73*). The γ-resorcylate catabolic pathway has been identified in *Rhizobium* sp. MTP-10005, where the gene for γRSD is downstream from genes encoding flavin reductase and resorcinol hydroxylase and upstream from genes encoding maleylacetate reductase and hydroxyquinol 1,2- dioxygenase as shown in **Scheme 4.2** (*74*). As compared to the organisms harboring LigW and LigW2, *Rhizobium* sp. MTP-10005 is also a gram-negative soil alphaproteobacteria, which can utilize various aromatic hydrocarbons as sole carbon and energy sources (*75*). Typically decarboxylases utilize



Scheme 4.1: Reaction catalyzed by γ-resorcylate decarboxylase (RSD).

an organic cofactor, such as pyridoxal 5'-phosphate or biotin or an inorganic cofactor, such as iron or zinc (*76*). There are also decarboxylases that function independent of any cofactors as shown in the case of orotidine monophosphate decarboxylase (OMPDC) (*77)*, *p*-coumarate decarboxylase (*78*), and methylmalonyl CoA decarboxylase (*79*). In comparison, LigW and γRSD (*24*) are members of COG2159 of the amidohydrolase superfamily, and the former requires a single divalent metal for catalysis. γRSD from *Rhizobium* sp. MTP-10005 has been characterized and shown to



Scheme 4.1: Catabolic pathway of γ-resorcylate in *Rhizobium* sp. MTP-10005. A, B, C, and D are represent γ-RSD, resorcinol hydrolase, hydroxyquinol 1, 2-diaoxygenase, and maleylacetate reductase, resopectively.

be a reversible decarboxylase with reported kinetic parameters k_{cat}, K_m, k_{cat}/K_m for the forward reaction of 0.2 s^{-1}, 71 µM, and 1.3 x 10^4 $M^{-1}s^{-1}$ and for the reverse reaction of 1.1 s^{-1}, 46 mM, and 98 $M^{-1}s^{-1}$, respectively (*60*). The reversibility of γRSD reaction has also been demonstrated using the enzyme from *Rhizobium* sp. MTP-10005 (*80, 81*).

In addition, the hypothetical protein (locus tag: Atu2529) isolated from *Agrobacterium tumefaciens* IAM1248 shares 96% sequence identity with γRSD from *Rhizobium* sp. MTP-10005. This enzyme also catalyzes the reversible carboxylation of resorcinol in addition to the decarboxylation of γ-resorcylate and 2, 3-dihydroxybenzoate (*17, 75*). The genes associated with the degradation of γ-resorcylate in *Agrobacterium tumefaciens* IAM1248 are presented in **Figure 4.1**. The reported activity of the purified

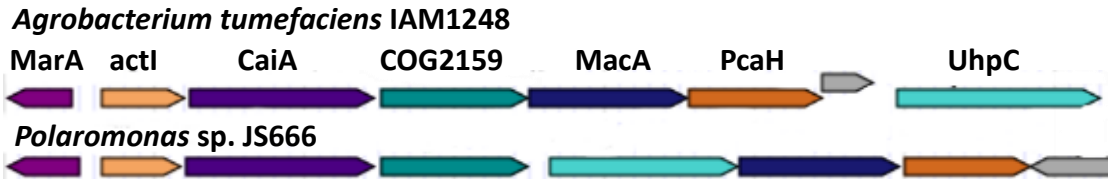


Figure 4.1: Genes related to γ-resorcylate catabolism seen in *Agrobacterium tumefaciens* IAM1248 and *Polaromonas* sp. JS666. Each color depicts a conserved gene within the operons of the two organisms. MarA, actI, CaiA, COG2159, MacA, PcaH , and UhpC correspond to transcriptional regulator, oxygenase component of resorcinol hydrolase, reductase component of resorcinol hydrolase, γRSD, maleylacetate reductase, hydroxyquinol 1, 2-dioxygenase, and benzoate transport.

γRSD from *Agrobacterium tumefaciens* IAM1248 is not dependent on an organic or metal cofactors and the specific activity for the conversion of γ-resorcylate to resorcinol was calculated to be 0.37 units/mg or 0.2 s^{-1} in the presence of 20 mM γ-resorcylate.

Furthermore, γRSD from Agrobacterium tumefaciens was reported to be stable in the pH range of 5-9. The K_m values for the decarboxylation of γ-resorcylate and 2, 3-dihydroxybenzoate were reported to be 0.08 mM and 0.12 mM, respectively. In contrast, the K_m value for the reverse reaction, the carboxylation of resorcinol (1, 3-dihydroxybenzene), was calculated to be 26 mM (*17*). γ-Resorcylate is used as an intermediate in the production of various chemicals by the medicinal and agricultural industries (*82*). The production of γ-resorcylate is prepared from resorcinol via Kolbe-Schmitt carboxylation, a process that is not regiospecific and requires both high temperatures and high pressures (*80*). Thus, the study of reversible decarboxylations carried out by various non-oxidative decarboxylases is of interest. Here we report the activity, metal dependency, substrate specificity, and crystal structure of γRSD from *Polaromonas* sp. JS666, complexed with 2-nitroresorcinol. This enzyme was previously annotated as amidohydrolase 2 (locus tag: Bpro_2061, gi| 91787937, EFI target: 500288) and it shares 77% sequence identity to γRSD from *Agrobacterium tumefaciens* IAM1248. The catabolic pathway for degradation of γ-resorcylate has been identified in this organism as illustrated **Figure 4.1**.

To aid in the search of reversible decarboxylases within COG2159, a hypothetical protein (locus tag: SAV2580, gi| 15925570, EFI target: 505312) from

Staphylococcus aureus subsp. aureus Mu50 was chosen to further the understanding of the enzymatic functions in COG2159. SAV2580 shares 32% and 29% sequence identity with γRSD and LigW, respectively. To analyze the homologous relationship between these proteins, a sequence similarity network (SSN) of COG2159 is generated at a BLAST E-value of 10^{-70} as shown in **Figure 4.2**.

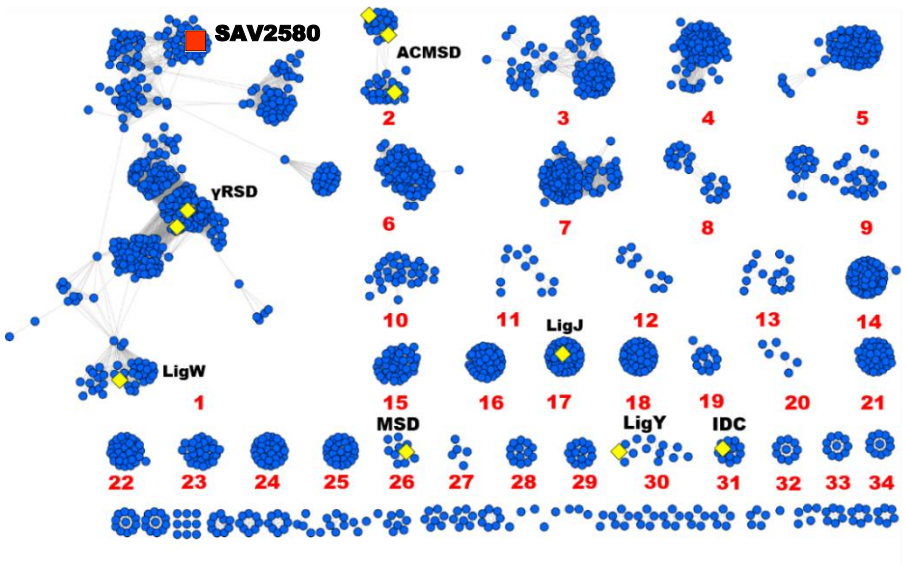


Figure 4.2: Sequence similarity network (SSN) generated by the software Cytoscape of COG2159 at a BLAST E-value of 10^{-70}. Group 1 contains γRSD, LigW, and the hypothetical decarboxylase SAV2180 from *Staphylococcus aureus subsp. aureus* Mu50 depicted as red square.

At BLAST E-value of 10^{-70}, group 1 contains SAV2180, LigW, and γRSD. It was determined that SAV2580 shares 34% sequence identity with a hypothetical decarboxylase (locus tag: BAV2745) found in *Bordetella avium* 197N. This enzyme in turn shares 38% and 37% sequence identity with 2, 3-dihydroxybenzoic acid decarboxylase (DHBD) from *Aspergillus oryzae* (*83*) and salicylate decarboxylase (Sdc) from *Trichosporon moniliiforme* WU-0401 (*84*), respectively. SAV2580, BAV2745, and DHBD are within group 1 with LigW and γRSD as shown in the SSN depicted in **Figure 4.3A**; however, all the proteins are separated into different sub-clusters. When the stringency is increased to BLAST E-value of 10^{-80} (**Figure 4.3B**), group 1 becomes 3 different sub-groups. Interestingly, γRSD and DHBD remain in sub-group 1.1 with 43% sequence identity. In addition to determining the substrate specificity of γRSD, this chapter describes the efforts towards determining the function of SAV2580.

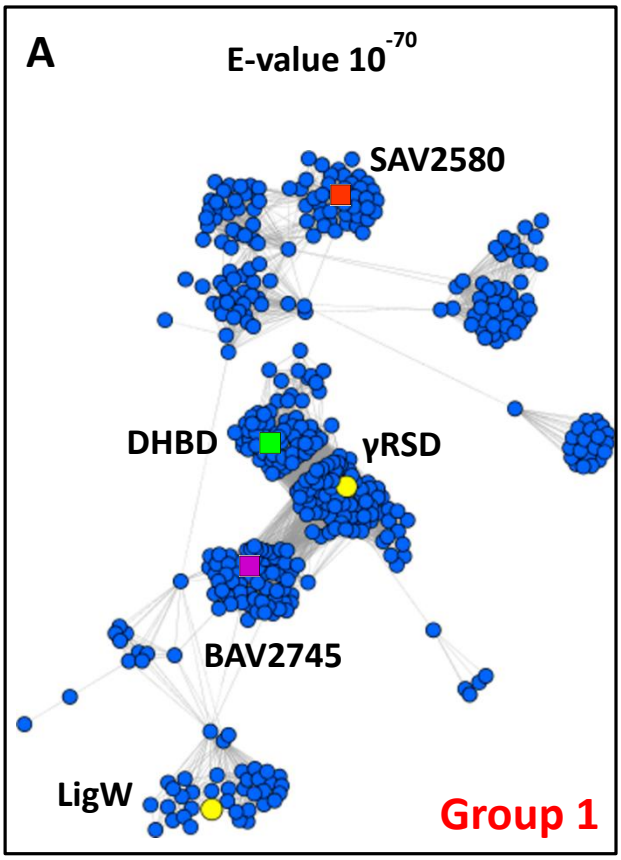


Figure 4.3: Sequence similarity network of group 1 in COG2159 at BLAST E-value of 10^{-70} (**A**) and 10^{-80} (**B**). Yellow circles represent the already characterized LigW and γRSD; whereas the green, red, and purple squares represent the 2, 3-dihydroxybenzoic acid (DHBD) and the hypothetical decarboxylases SAV2580 and BAV2745, respectively. At BLAST E-Value of 10^{-80}, group one is separated into sub-clusters, where 1.1 contains γRSD and DHBD and 1.2, 1.3, and 1.4 encompass LigW, BAV2745, and SAV2580, respectively.

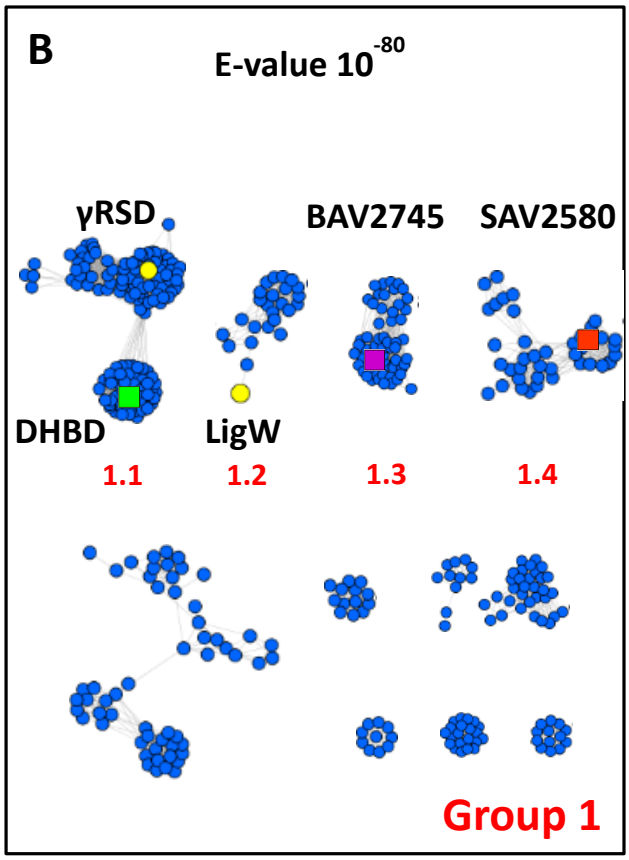


Figure 4.3: Continued.

MATERIALS AND METHODS

Materials. All chemicals, buffers, and purification reagents were purchased from Sigma-Aldrich unless otherwise specified. *Pfu Turbo* DNA polymerase and the *E. coli* strains BL21 (DE3) and XL1-blue cells were obtained from Stratagene. The restriction enzyme DpnI was purchased from New England BioLabs and oligonucleotides were obtained from Integrated Data Technology through the Gene Technology Laboratory at Texas A&M University. Kanamycin, isopropyl β-D-thiogalactopyranoside (IPTG), and LB broth were acquired from Research Products International Corp. Protamine sulfate, Wizard Plus SV Miniprep DNA purification system, and Vivaspin centrifugal concentrator (10 000 MWCO) were obtained from MP Biomedicals LLC., Promega, and Fisher Scientific, respectively. Chromatographic gel filtration columns were purchased from GE Healthcare.

Expression and Purification of γ-Resorcylate Decarboxylase (γRSD) from *Polaromonas* sp. JS666. The recombinant plasmid containing the γRSD gene (Bpro_2061; gi: 91787937; EFI target: 500288) from *Polaromonas* sp. JS666 was obtained from the laboratory of Dr. Steve Almo at the Albert Einstein School of Medicine through the Enzyme Function Initiative (EFI). The clone encoded codon-optimized the γRSD with a C-terminal Met-Val insertion and N-terminal Ala-Glu-Asn-Leu-Tyr-Phe-Gln-Ser-His_6 insertion. The N-terminal His-tag was cleaved via standard site-directed mutagenesis protocols using the following primer 5'-GAGAACCTCTACTTCTAATAGCACCATCATCACCACCAT-3' (forward) and 5'-ATGGTGGTGATGATGGTGCTATTAGAAGTAGAGGTTCTC-3' (reverse).

The recombinant plasmid was transformed by electroporation into BL21 (DE3) cells and plated on agar plates containing 50 µg/mL filtered kanamycin. Six colonies were inoculated into 7 mL LB broth containing 7 µg/mL kanamycin and grown overnight at 37 °C. The starting cultures were used to inoculate a total of 6 L LB broth containing 50 µg/mL kanamycin and left to incubate at 37 ^{0}C by shaking. Isopropyl β-thiogalactoside (IPTG) was added to a final concentration of 0.5 mM to each of the flasks once the absorbance reached a range between 0.4-0.6 OD. Upon induction, 2 L of cells were supplemented with 1 mM $MnCl_2$ and another 2 L with 1 mM $ZnCl_2$. The two flasks remaining were not supplemented with any metals at the time of induction. Expression of the protein was achieved by shaking in 22 °C overnight. The cells were then collected by centrifugation (8000 rpm, 10 min) and stored in the -80 °C freezer.

The frozen cells were thawed at room temperature and suspended in 20 mM HEPES (pH 7.5) containing 20 mg phenylmethanesulfonyl fluoride (PMSF) for every 5 g of cells. Lysis was achieved by sonication, followed by centrifugation at 8000 rpm for 10 min, and addition of 0.1 g of protamine sulfate dissolved in 10 mL of 20 mM HEPES (pH 7.5) by 45 minute drip to the supernatant. The solution was then centrifuged at 8000 rpm for 10 minutes and 13.3 g ammonium sulfate was added to the supernatant to achieve 45% saturation. After stirring for 45 minutes, the suspension was centrifuged at 8000 rpm for 10 minutes. The pellet was then re-suspended in 10 mL of 20 mM HEPES, pH 7.5 and loaded onto a 26/60 Superdex 200 gel-filtration column. The protein was eluted using 20 mM HEPES pH 7.5 and further loaded on a Resource-Q (GE Healthcare) anion-exchange column (6 mL). γRSD was eluted using a gradient of 1 M NaCl in 20

mM HEPES pH 7.5. The protein was concentrated using Vivaspin 20 for 10 kDa MW obtained from GE Healthcare and stored in the -80 ^{o}C freezer.

Kinetic Parameters. The decarboxylations of γ-resorcylate (γRS) by γRSD were determined by monitoring the decrease in absorbance upon formation of resorcinol at 305 nm using a differential molar extinction coefficient ($\Delta\varepsilon_{305}$) of 3,208 M^{-1} cm^{-1}. Alternatively, the decarboxylation reactions of 2, 3- dihydroxybenzoic acid, 2, 4, 6-trihydroxybenzoic acid, and 2, 6-dihydroxy-4-methylbenzoic acid were monitored by following the decrease in absorbance at 306, 256, and 304 nm using the differential extinction coefficients ($\Delta\varepsilon_{306}$) of 3680 M^{-1} cm^{-1}, ($\Delta\varepsilon_{256}$) of 5953 M^{-1} cm^{-1}, and ($\Delta\varepsilon_{304}$) of 2977 M^{-1} cm^{-1}, respectively. All assays were conducted using a 96-well SpectraMax Plus384 UV-vis spectrophotometer and the steady-state kinetic parameters were obtained using SoftMax Pro 5.0. Standard assay conditions for the decarboxylation of γRS, 2, 3- dihydroxybenzoic acid, 2, 4, 6-trihydroxybenzoic acid, and 2, 6-dihydroxy-4-methylbenzoic acid included 50 mM HEPES at pH 7.0, substrate range of 0-500 μM, and 0.1 μM γRSD in a final volume of 250 μL at 30 °C. The enzyme activities of SAV2580 and γRSD were probed using a library of hydroxybenzoate compounds (**see Appendix**) by comparing the UV-vis spectra in the range of 220 to 400 nm of the reaction mixture and a control without enzyme. Typically, the reactions were carried out with in 50 mM HEPES at pH 7.0, and with 2 μM enzyme. The decarboxylation of 1-hydroxy-2-naphthanoic acid by SAV2580 was determined by monitoring the decrease in absorbance upon formation of 1-napthol at 340 nm using a differential molar extinction coefficient ($\Delta\varepsilon_{340}$) of 4275 M^{-1} cm^{-1}. All kinetic constants were obtained using equation

4.1, where v is the initial velocity, E_t is the total enzyme concentration, [A] is the substrate concentration, k_{cat} is the turnover number, and K_m is the Michaelis constant.

$$v/E_t = k_{cat}\,[A] / (K_m + [A]) \qquad (4.1)$$

Carboxylation of Resorcinol by γRSD. The carboxylation of resorcinol was carried out in an air tight vial by bubbling CO_2 for 30 min. The reaction mixture contained 30 mM resorcinol, 50 mM phosphate buffer (pH 7.0), 1 mM $ZnCl_2$, 638 mM $NaHCO_3$, and 3.5 μM γRSD. The UV-Vis spectra in an identical buffer systems containing resorcinol in absence of γRSD, γ-resorcylate in absence of γRSD, and the reaction of γRSD with resorcinol after 30 min and 12 hours were compared in the range of 220 – 400 nm.

Metal Analysis. The metal content of γRSD was determined by inductively coupled plasma mass spectrometry (ICP-MS) with an Elan DRC II ICP-MS instrument from Perkin-Elmer. Prior to ICP-MS analysis, loosely bound metals were removed by passage through a PD-10 desalting column equilibrated with 50 mM HEPES (pH 7.0) previously treated with Chelex 100 ion exchange resin. The protein sample was treated with concentrated HNO_3 for 15 minutes at 100 °C and then diluted with distilled water to a final protein concentration of 1.0 μM and 1% (v/v) HNO_3.

The activation of γRSD was probed using Mn^{2+}, Zn^{2+}, and Co^{2+} by adding these metals (0 - 3 mM) to the buffered reaction mixture. The activity of γRSD was also probed in presence of 1 mM of the following metals, Mg^{2+}, Zn^{2+}, Co^{2+}, Mn^{2+}, Ca^{2+}, Na^{+}, and K^{+}, respectively in 50 mM HEPES (pH 7.0), 480 μM γ-resorcylate, and 2 μM γRSD. The activity was measured at 30 °C immediately after addition of these metals to the reaction mixture. γRSD was overexpressed without the addition of metal to the growth

medium and this enzyme was incubated with 10 mM o-phenanthroline and EDTA, respectively at 4 °C in 50 mM MES at pH 6.5 for 12 hours. A PD-10 column, previously equilibrated with 50 mM MES at pH 6.5, was used to remove the o-phenanthroline or EDTA prior to addition of 2 or 0.2 mM $ZnCl_2$ or $mMnCl_2$ to the reaction buffer.

Activity Measurement of 505312 at Different pH-Rate Values. The kinetic constants (k_{cat}, K_m, and k_{cat}/K_m) for the conversion of 1-hydroxy-2-naphthoic acid to 1-naphthol by SAV2580, were determined at pH 7.0 in 50 mM HEPES, 1 μM enzyme, and 0-450 μM 1-hydroxy-2-naphthoic acid using the differential extinction coefficient $\Delta\varepsilon_{340}$ of 4275 M^{-1} cm^{-1}.

Inhibition Studies. 2-Nitroresorcinol (2-NR) was utilized as an inhibitor for the reaction catalyzed by γRSD. The inhibition experiments for 2-NR contained 60 μM γ-resorcylate, 50 mM HEPES (pH 7.0), and 1.0 μM γ-RSD in a total volume of 250 μL at 30 °C. The apparent inhibition constant (K_i^{app}) for 2-NR was obtained from a fit of the data to equation 4.2, where [I] is the inhibitor concentration, v_0 is the velocity in absence of inhibitor, and v_i is the velocity in presence of inhibitior. The K_d value was obtained from equation 4.3, where [S] is the substrate concentration used in the inhibition experiment and K_m is the Michaelis constant for the substrate γRSD.

$$v_i/v_\circ = K_i^{app} /([I] + K_i^{app}) \quad (4.2)$$

$$K_i^{app} = K_d (1 + [S]/K_m) \quad (4.3)$$

Crystallization of γRSD from *Polaromonas* sp. JS666. The high resolution crystal structures of γRSD from *Polaromonas* sp. JS666 in the presence of 2-nitroresorcinol (2-NR) was solved in collaboration with the laboratory of Dr. Steven C.

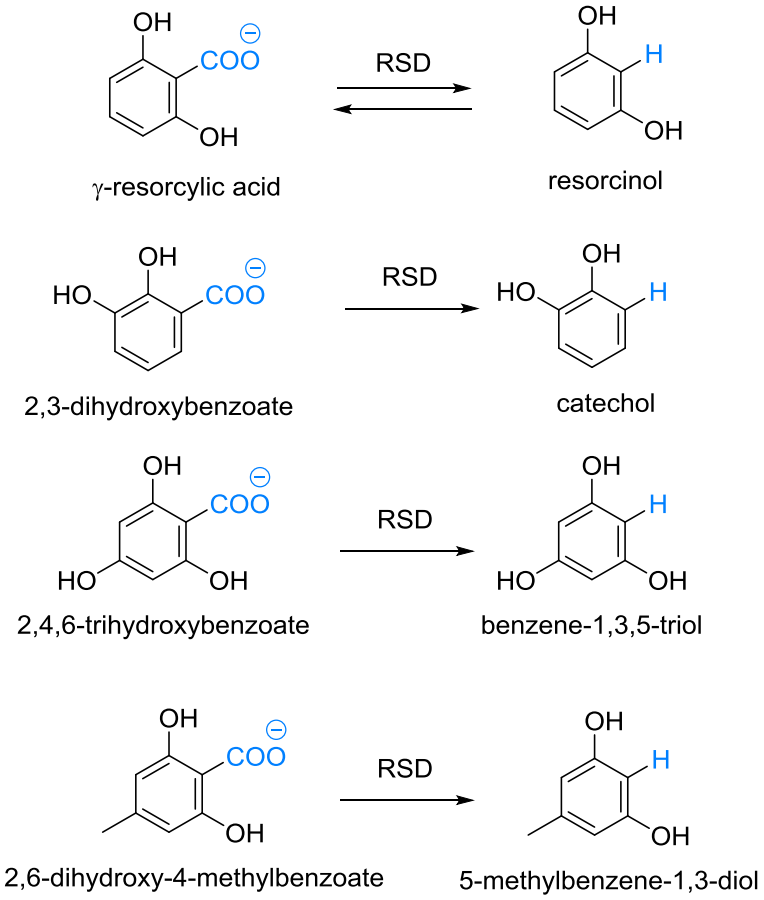
Almo at the Department of Biochemistry, Albert Einstein College of Medicine, at Bronx, New York. The PDB code for the Mn^{2+} substituted γRSD in presence of 2-NR is 4QRO.

RESULTS

Purification and Properties of γRSD from *Polaromonas* sp. JS666. γRSD converts γ-resorcylate to recorcinol, 2,3-dihydroxybenzoic acid to catechol, 2, 4, 6-trihydroxybenzoic acid to benzene-1,3,5-triol, and 2,6- dihydroxy-4-methylbenzoic acid to 5-methylbenzene-1,3-diol as shown in **Scheme 4.3**. γRSD was purified to homogeneity and the as-purified enzyme exhibited catalytic activity for the decarboxylation of γ-resorcylate to resorcinol (**Figure 4.3A**) with values of k_{cat} = 0.44 ± 0.01 s^{-1} and k_{cat}/K_m = 1.3 x 10^4 M^{-1} s^{-1}. The catalytic constants for the decarboxylation reaction carried out by γRSD are shown in **Table 4.1**, the conversion of 2, 3-dihydroxybenzoic acid, 2, 4, 6-trihydroxybenzoic acid, and 2, 6-dihydroxy-4-methylbenzoic acid were monitored with decrease of absorbance at 306, 256, and 304 nm (**Figure 4.3B, C, and D**).

Table 4.1: Kinetic constants of γRSD

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
γ-resorcylate	0.44 ± 0.01	32 ± 2.1	1.3 x 10^4
2,3-dihydroxybenzoate	0.47 ± 0.02	111 ± 11.1	4.2 x 10^3
2,4,6-trihydroxybenzoate	0.18 ± 0.01	48 ± 1.3	3.8 x 10^3
2,6-dihydroxy-4-methylbenzoate	0.38 ± 0.01	79 ± 5.8	4.8 x 10^3



Scheme 4.3: Reactions catalyzed by γ-resorcylate decarboxylase (γRSD) from *Polaromonas* sp. JS666.

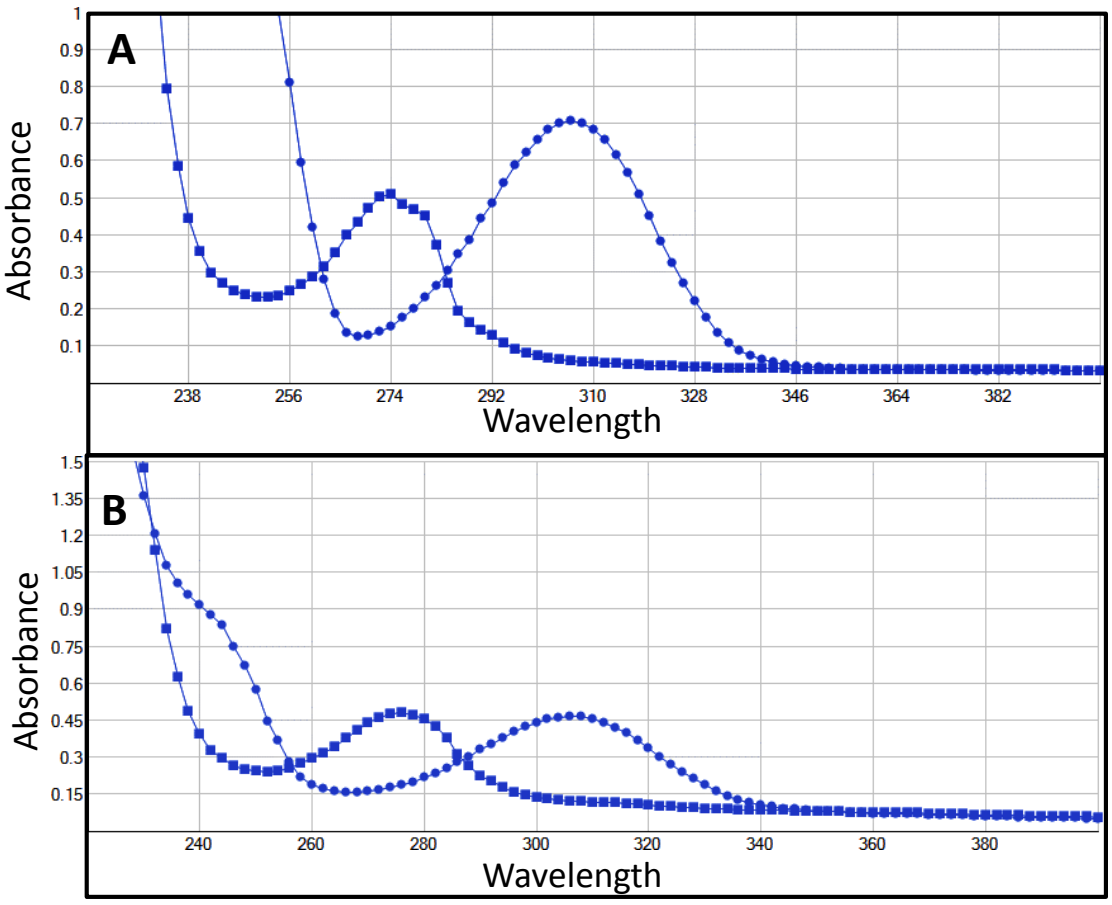


Figure 4.4: UV-Vis spectra of γ-resorcylate and recorcinol (**A**), 2,3-dihydroxybenzoic acid and catechol (**B**), 2, 4, 6- trihydroxybenzoic acid and benzene-1,3,5-triol (**C**), and 2,6-dihydroxy-4-methylbenzoic acid and 5-methylbenzene-1,3-diol (**D**). In each spectrum, the substrates γ-resorcylate, 2,3-dihydroxybenzoic acid, 2, 4, 6- trihydroxybenzoic acid, and 2,6-dihydroxy-4-methylbenzoic acid are represented as blue circles; whereas, the products recorcinol, catechol, 5-methylbenzene-1,3-diol, and 2,6-dihydroxy-4-methylbenzoic acid are represented as blue squares.

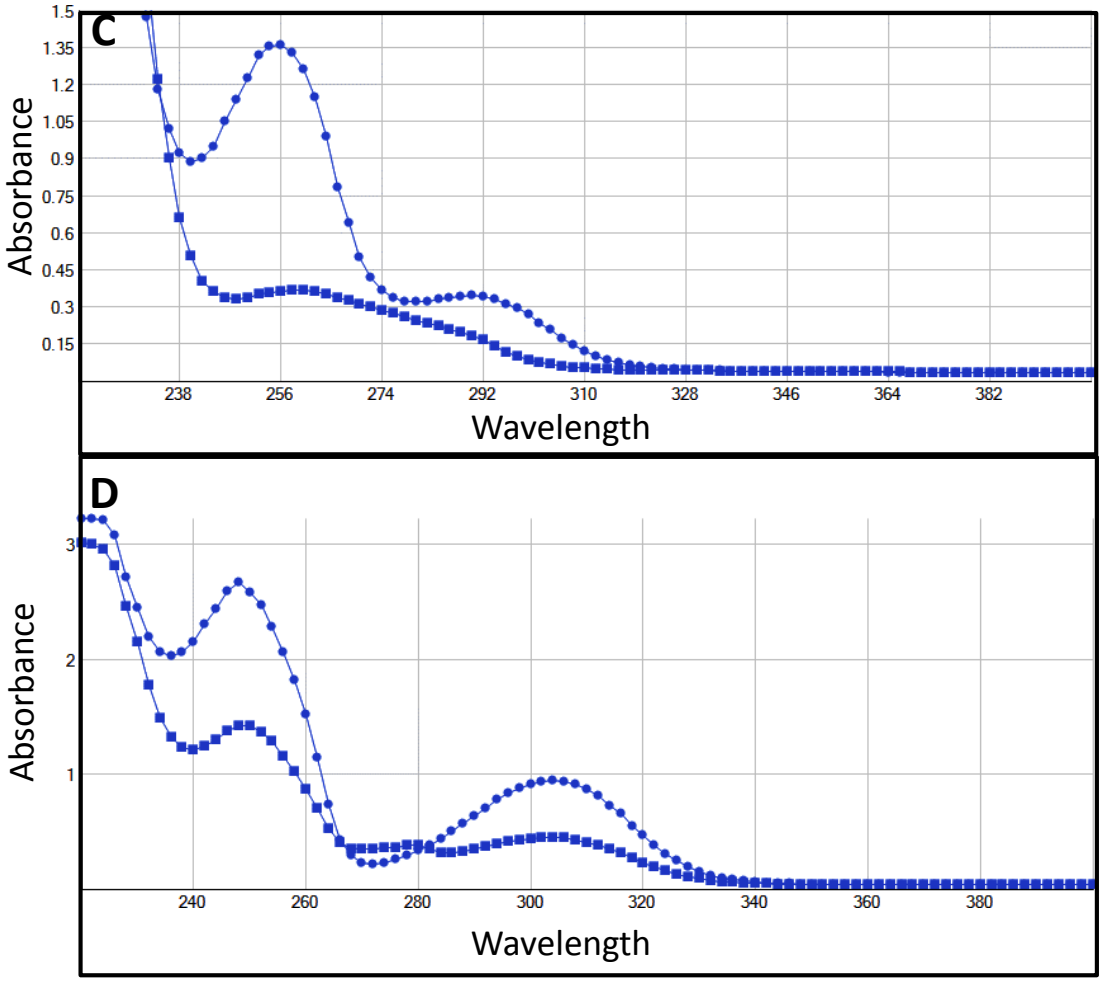


Figure 4.4: Continued.

Metal Content Analysis. The as purified γRSD contained 0.04 and 0.05 equivalents of Fe^{2+} and Ni^{+2} in the active site as determined by inductively coupled plasma mass spectrometry (ICP-MS). Improvement of activity was tested by adding various 1 mM metal salts, including $MgCl_2$, $ZnCl_2$, $CoCl_2$, $MnCl_2$, $CaCl_2$, NaCl, and KCl to the reaction mixture (**Figure 4.5**). The kinetic constants (k_{cat} and k_{cat}/K_m) for the decarboxylation of γ-resorcylate by γRSD were lowered by approximately three fold, when the enzyme was incubated with 2 mM *o*-phenanthroline or EDTA, respectively at pH 7.0 for 12 hrs at 4 °C. When Zn^{2+} or Mn^{2+} were added directly to the reaction mixture containing the chelator treated enzyme. In presence of 0.2 mM Mn^{2+} the activity of γRSD was reconstitution back to the wild-type activity; whereas, no effect was observed by the presence of 0.2 mM Zn^{2+} (**Table 4.2**).

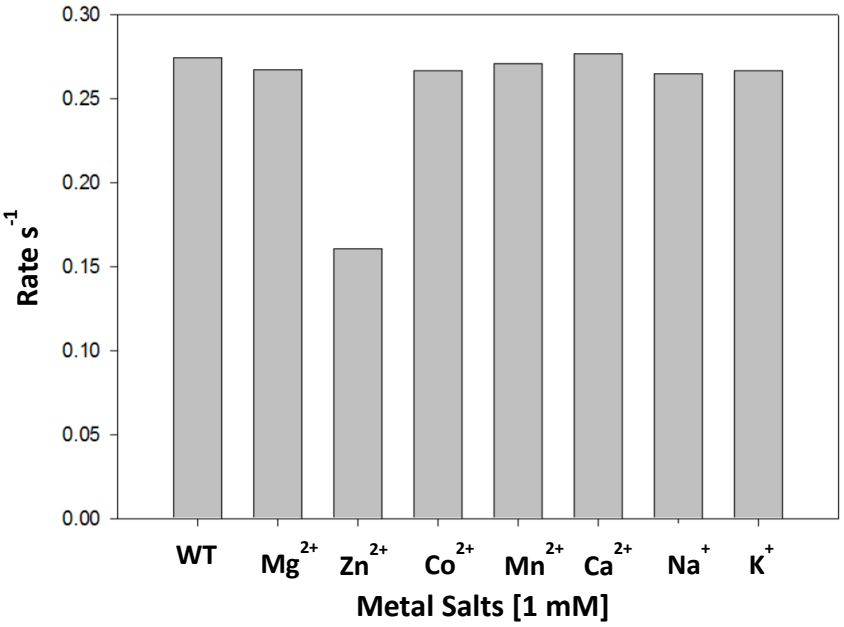


Figure 4.5: Activity of γRSD in presence of various metal chlorides.

Therefore, γRSD was subsequently expressed in *E. coli* in the presence of 1.0 mM $MnCl_2$ or 1.0 mM $ZnCl_2$ that were added directly to the growth medium. The metal content of γRSD overexpressed in the presence of $MnCl_2$ and $ZnCl_2$ and further purified with 50 mM HEPES (pH 7.0) contained an average of 0.2 equivalents of Mn^{2+} and 0.06 equivalents of Zn^{2+} per monomer with catalytic constants of k_{cat} = 0.362 ± 0.004 s^{-1} and k_{cat}/K_m = 1.0 x 10^4 M^{-1} s^{-1} and k_{cat} = 0.32 ± 0.01 s^{-1} and k_{cat}/K_m = 9.6 x 10^3 M^{-1} s^{-1}, respectively. The values for γRSD obtained under Mn^{2+} and Zn^{2+} expression conditions are close to the kinetic constants of the wild-type enzyme that was expressed in the absence of added metal to the growth medium.

Table 4.2: Kinetic constants of γRSD for the conversion of γ-resorcylate to resorcinol treated with a chelator and subsequently with various metals

γRSD overexpressed in absence of metal	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
WT γRSD	0.44 ± 0.01	32 ± 2.1	1.3 x 10^4
Incubated EDTA	0.16 ± 0.01	30 ± 4.5	5.3 x 10^3
Incubated o-phen.	0.12 ± 0.01	27 ± 4.4	4.4 x 10^3
Addition of 0.2 mM Zn^{2+} after removal of EDTA	0.070 ± 0.003	17 ± 2.2	4.0 x 10^3
Addition of 0.2 mM Zn^{2+} after removal of o-phen.	0.081 ± 0.011	32 ± 11	2.5 x 10^3
Addition of 0.2 mM Mn^{2+} after removal of EDTA	0.203 ± 0.007	57 ± 7.1	3.5 x 10^3
Addition of 0.2 mM Mn^{2+} after removal of o-phen.	0.49 ± 0.10	34 ± 18	1.4 x 10^4

Inhibition and Crystal Structure of γRSD. 2-nitroresorcinol (2-NR) was selected as a potential inhibitor of the reaction catalyzed by γRSD. The apparent inhibition constant (K_i^{app}) of 96 ± 8 μM was obtained at a substrate concentration of 60 μM (**Figure 4.6**). The dissociation constant for this compound was determined to be 40 μM after correcting for the concentration of the substrate used in the experiments. The high resolution crystal structure (1.65 Å) of the wild-type γRSD from *Polaromonas* sp. JS666 complex with manganese and 2-nitroresorcinol (2-NR, PDB id: 4QRO), shows that the metal in the active site has an octahedral geometry (**Figure 4.7**). The active site resembles closely the active sites of the Mn^{2+} substituted LigW and LigW2 complexed

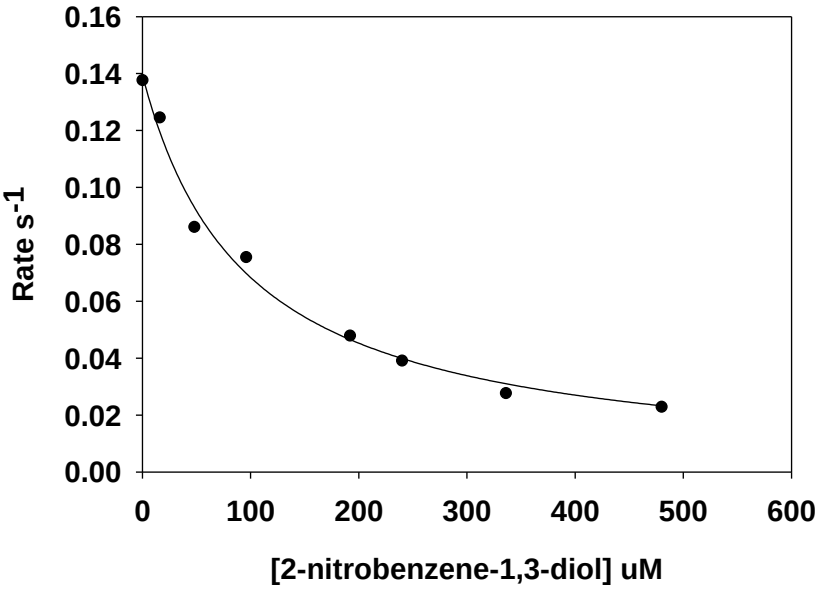


Figure 4.6: Inhibition of the reaction catalyzed by γRSD in the presence of 2-nitroresorcinol (2-NR) with data fitted to equation 4.2, where K_i^{app} is the slope inhibition constant.

with 5-nitrovanillate (5-NV, PDB id: 4NG3 and 4QRN, respectively). In γRSD, the ligand 2-NR displaying the bidentate binding very similar to that observed in the LigW and LigW2 structures (PDB id: 4NG3 and 4QRN). In addition, the residues important for binding the Mn^{2+} ion include a glutamate and a histidine from the ExH motif found in β-strand 1 (Glu-8 and His-10), a histidine from β-strand 5 (His-164), and an aspartate from β-strand 8 (Asp-287) as illustrated in **Figure 4.7**. Also, the overall topology of γRSD resembles the close homologues LigW, LigW2, ACMSD, and IDC. The ribbon representation of γRSD is shown in **Figure 4.8A**, indicating Mn^{2+} substituted γRSD as a homotetramer, which is consistent with the structure of γ-resorcylate decarboxylase (γRSD) from *Rhizobium* sp. (PDB id: 2DVU).

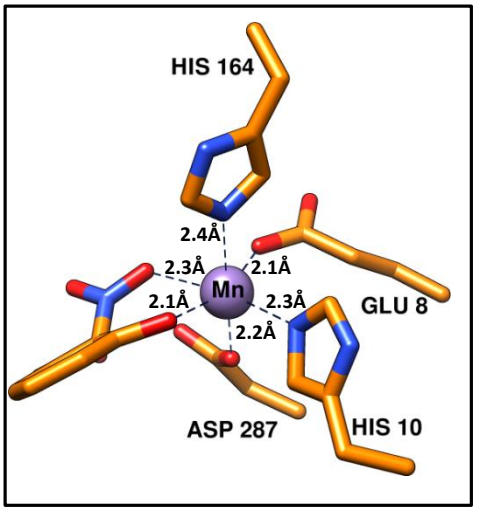


Figure 4.7: Metal coordination seen in the structure of γRSD from *Polaromonas* sp. JS666 (PDB id: 4QRN) in presence of Mn^{2+} and 2-nitroresorcinol (2-NR).

Although γRSD from *Polaromonas* sp. JS666 and γRSD from *Rhizobium* sp. share 33% sequence identity, the overall active site architecture is conserved; however, the binding of the nitro inhibitor differs from the positioning of γ-resorcylate and metal ligation seen in the substrate bound γRSD structure form *Rhizobium* sp. (PDB id: 2DVU, **Figure 4.9A**). In PDB id: 4QRO, each protein subunit consists of distorted $(\beta/\alpha)_8$ -TIM barrel (**Figure 4.8B**) domain with the active site and Mn^{2+} ion embedded towards the C-terminus of the barrel. An insertion domain is observed in γRSD from *Polaromonas* sp. JS666 containing α-helix as observed in γRSD from *Rhizobium* sp. This insertion domain contains Phe-23, which participates in a T-shaped stacking with the phenyl ring of the nitro inhibitor with a closest van der Waals distance of 3.8Å. The

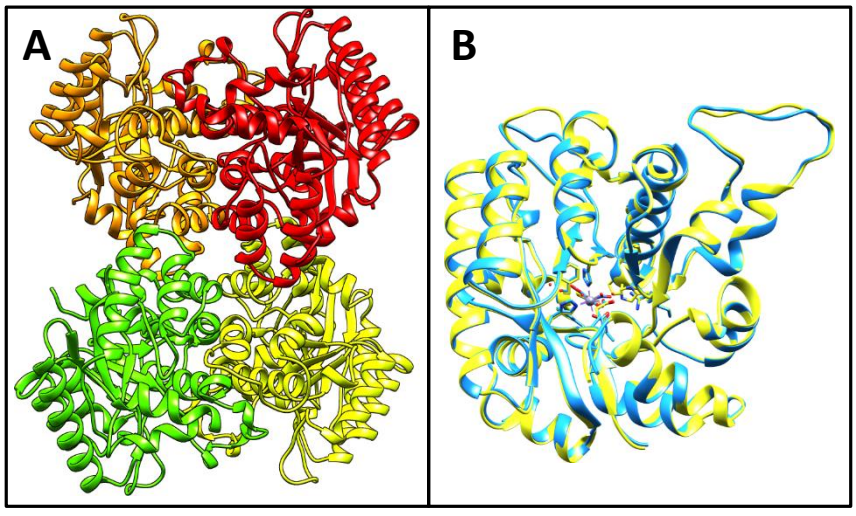


Figure 4.8: Quaternary structure of in γRSD from *Polaromonas* sp. JS666 (**A**) and the monomeric superposition of γRSD from Polaromonas sp. JS666 (PDB id: 4QRO) in yellow with γRSD from *Rhizobium* sp. (PDB id: 2DVU) in cyan (**B**).

inhibitor is further stabilized through π-π stacking with Phe-290 with a distance of 3.5Å. Similar to LigW and LigW2, two conserved residues from β-strand 6, a histidine (His-218) and a glutamate (Glu-221) two residues apart (HxxE), form a hydrogen bond triad together with Asp-287 from β-strand 8 in γRSD **Figure 4.9B**. In addition, Arg-229 from the adjacent monomer is within 3.2Å from one of the oxygen atoms of the C-2 nitro group. In contrast, Asn-234 from the same subunit in turn hydrogen bonds to the C-6 hydroxyl group as observed in the crystal structure of the substrate bound γRSD from *Rhizobium* sp. However, the side chain of this amino acid is flipped in the γRSD from *Polaromonas* sp. (PDB id: 4QRO) crystal structure, so that the nitrogen of the carboxamide is in place of the oxygen atom, when compared to the structure of γRSD from *Rhizobium* sp. with γ-resorcylate (PDB id: 2DVU). In the latter crystal structure, the substrate is postioned so that the C-2 hydroxyl group hydrogen bonds to a water molecular, which in turn is hydrogen bonded to Ser-20, this interaction is not observed in nitro inhibitor structure (PDB id: 4QRO), since the C-2 hydroxyl group is metal ligated (**Figure 4.9**). Furthermore, it is observed that the nitro inhibitor geometry deviates from planarity, in particular the C-2 nitro group is tilted from the main phenyl ring plane with a torsion angle of 51° between the N, C-2, C-1, and the oxygen of the C-1 hydroxyl group; whereas, in the crystal structure of γRSD (PDB id: 2DVU) with the substrate γ-resorcylate the torsion angle between the carbon of the C-2 carboxyl group, C-2, C-1, and the C-1 hydroxyl oxygen is 0.15 making the molecule virually planer.

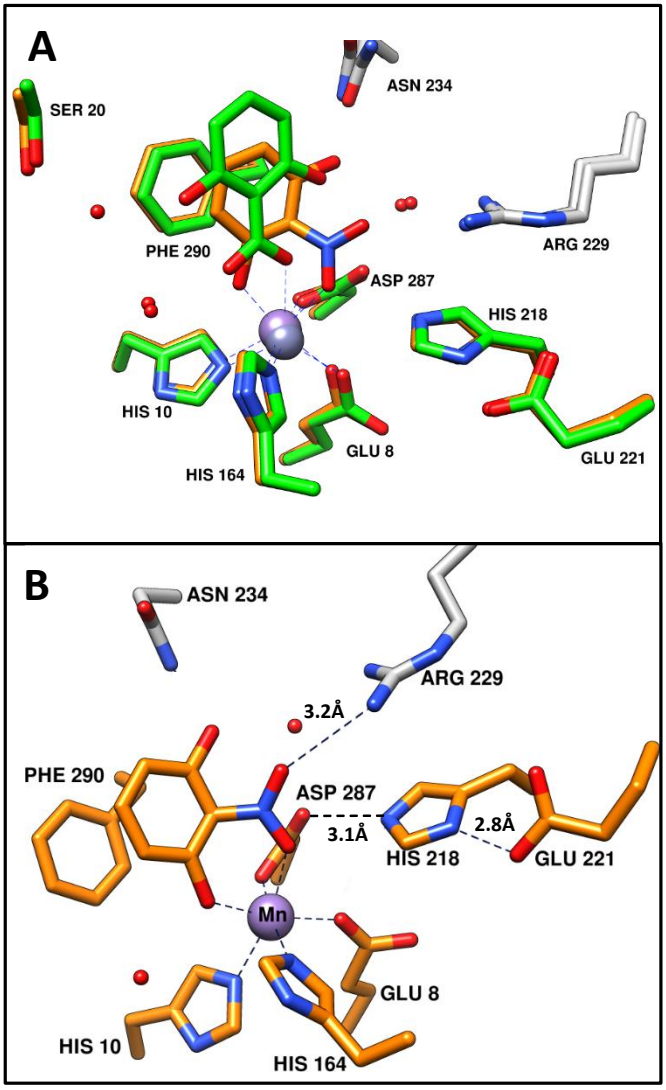


Figure 4.9: Active site overlay of in γRSD from *Polaromonas* sp. JS666 complexed with 2-NR and Mn^{2+} (PDB id: 4QRO) and γRSD from *Rhizobium* sp. complexed with γ-resorcylate and Zn^{2+} (PDB id: 2DVU) (**A**) The active site of PDB id: 4QRO (**B**).

Activity of SAV2580 from *Staphylococcus aureus subsp. aureus* Mu50.

SAV2580 catalyzes the conversion of 1-hydroxy-2-naphthoic acid to naphthalene-1-ol (**Figure 4.10A**) with catalytic constants k_{cat}, K_m, and k_{cat}/K_m of 0.09 ± 0.01 s^{-1}, 669 ± 109 μM, and 1000 ± 70 M^{-1} s^{-1}. The activity was monitored by following the decrease of absorbance upon conversion of 1-hydroxy-2-naphthoate to napththalen-1-ol at 340 nm (**Figure 4.1B**).

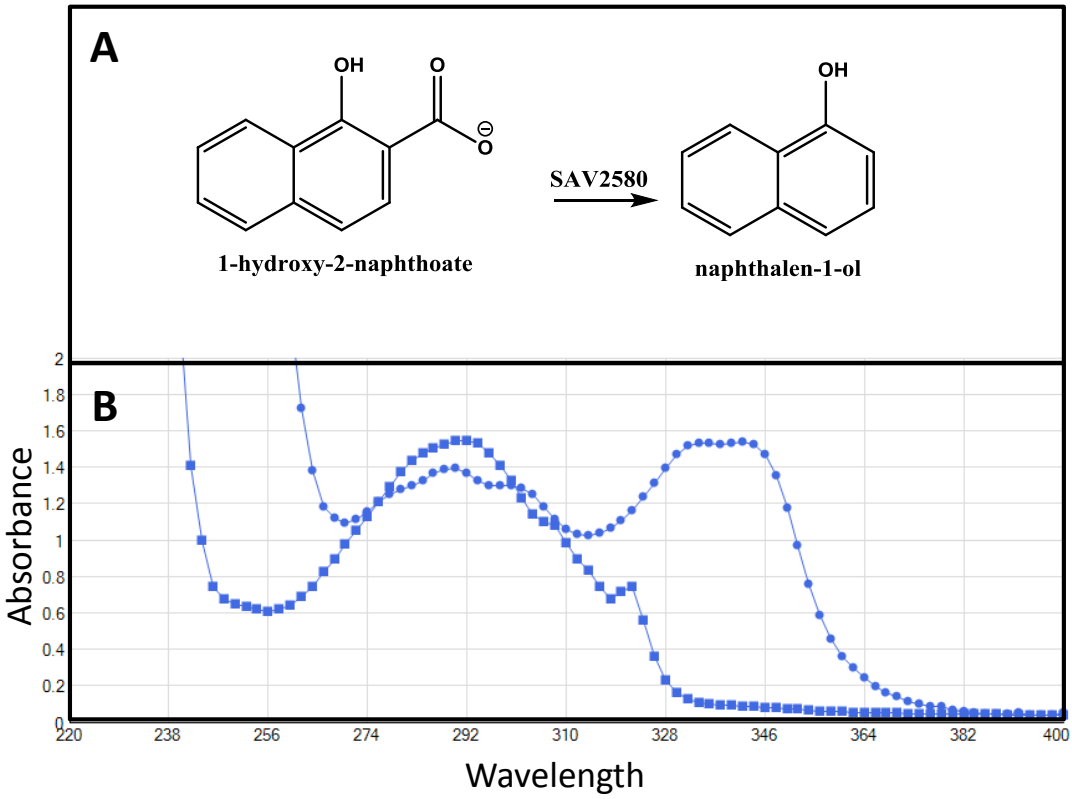


Figure 4.10: Reaction catalyzed by SAV2580 (**A**). UV-Vis spectra of 1-hydroxy-2-naphthoic acid in blue circles and naphthalene-1-ol in blue squares (**B**).

DISCUSSION

Structure Elucidation of γRSD and Substrate Promiscuity. The crystal structure of γRSD from *Polaromonas* sp. JS666 complexed with 2-NR and Mn^{2+} (PDB id: 4QRO) was solved to 1.65 Å. The enzyme fold as a distorted $(\beta/\alpha)_8$-barrel and a single divalent cation is bound in the active site, which is located at the C-terminus of the barrel. The active site residues are identical in both structures of γRSD, namely the enzyme isolated from *Rhizobium* sp. (PDB id: 2DVU) and γRSD from *Polaromonas* sp. JS666 (PDB id: 4QRO). The latter enzyme shares 33 % sequence identity with γRSD from *Rhizobium* sp., LigW, and LigW2, respectively. The quaternary structure of γRSD is a homotetramer, but the minimal catalytic state is a homodimer. An essential arginine residue conserved in all COG2159 decarboxylases from the adjacent subunit extends into the active site and presumably stabilizes the leaving carboxylate. The 1.65 Å structure of γRSD in presence of Mn^{2+} and the strong inhibitor 2-nitroresorcinol (2-NR) shows the active site metal center in an octahedral geometry retained by the bidentate ligation of one of the oxygen atoms of the C-2 nitro group and the oxygen of the C-1 hydroxyl group. Amino acids involved in the coordination of the divalent metal cation include an ExH motif from β-strand 1, a histidine from β-strand 5, and an aspartate from β-strand 8. The latter residues are highly conserved in all the COG2159 decarboxylases; whereas, the construction of the motif in β-strand 1 varies among the members. In addition the HxxE motif found in β-strand 6 is also conserved in LigW and LigW2 and similarly participates in a hydrogen bond triad. The hydrophobic residues Phe-23 and Phe-290 stabilize the phenyl ring of the substrate analogue through van der Waals and π-

π stacking interactions. In LigW these residues correspond to Tyr-51 and Tyr299. Thus, the substrates of LigW and γRSD are limited to having a planer geometry.

In addition, the distorted ligand 2-NR in γRSD is very similar to 2-nitrovanillate (2-NV) in LigW. The nitro group deviates from planarity in a similar fashion as 2-NV in LigW; however, the puckering observed at the aromatic carbon bearing the nitro group is only observed in the near angstrom structure of LigW2 (PDB id: 4QRN). Nevertheless, the bending of the nitro substituent (**Figure 4.7**) is unique in the γRSD structure with 2-NR (PDB id: 4QRN) in comparison to γRSD complexed with γ-resorcylate (PDB id: 2DVU), where the carboxylate group is not bent and the molecule remains planer (**Figure 4.9A**). Thus, 2DVU, may represent a non-productive complex between the enzyme and the substrate, γ-resorcylate. Furthermore, in 2DVU the substrate is further coordinated to the single zinc cation through one of the oxygen atoms of the leaving carboxylate group; thus, displaying a 5-coordinated geometry. Although the metal dependency for γRSD has not been completely resolved, the high similarity between the structures 4QRO (γRSD-Mn-2-NR) and 4QRN (ligW2-MN-5-NV) may suggest similar catalytic mechanism between the two enzymes.

Substrate Promiscuity of γRSD. In addition to carrying out the decarboxylation of γ-resorcylate to resorcinol, γRSD was previously shown to convert 2, 3-dihydroxybenzoate to catechol (*75*). In the sequence similarity network (SSN) of enzymes found in COG2159, γRSD remains clustered with 2, 3-dihydroxybenzoate decarboxylase (DHBD) isolated from *Aspergillus niger* at a high stringency of BLAST E-value of 10^{-70} (**Figure 4.4B**). DHBD is a non-oxidative decarboxylase functioning independent of an organic or inorganic cofactors and catalyzing the last step the fungal

metabolism of indole to catechol (*83*). γRSD and DHBD share 43% sequence identity; thus, DHBD may also catalyze the decarboxylation of γ-resorcylate. In the current study, it was determined that γRSD also carries out the decarboxylation of 2, 4, 6-trihydroxybenzoate to benzene-1, 3, 5-triol and 2, 6-dihydroxy -4 methylbenzoate to 5-methylbenzene-1, 3,-diol. The catalytic constants k_{cat} and k_{cat}/ K_m for the conversion of 2, 4, 6-trihydroxybenzoate are approximately two fold and three fold lower than the catalytic constants for the decarboxylation of γ-resorcylate. In contrast, the turnover number for the decarboxylation of 2, 6-dihydroxy-4-methylbenzoate is the same as the k_{cat} for γ-resorcylate; however, the catalytic efficiency for this substrate is also lowered by three fold.

Metal Analysis of γRSD. The catalytic turnover and efficiency for the decarboxylation of γ-resorcylate to resorcinol were not improved in presence of various metal chlorides in the reaction buffer. Incubation of as-purified γRSD obtained without addition of metal to the media contained less than 0.1 eq. metal in the active site, incubation of this enzyme with a metal chelator; however, decreased the catalytic constants by at least three fold for both k_{cat} and k_{cat}/K_m. Addition of 0.2 mM Mn^{2+} after removal of the chelator, reconstituted the original activity with k_{cat} and k_{cat}/K_m of 0.49 ± 0.10 s^{-1} and 1.4 x 10^4 M^{-1} s^{-1}. Further studies are necessary to determine the metal dependency of this enzyme.

Activity SAV2580 from *Staphylococcus aureus subsp. aureus* Mu50A. The hypothetical decarboxylase SAV2580 is found within the cluster of γRSD and LigW in the SSN at a BLAST E-Value of 10^{-70} and shares 33% and 29% sequence identity with γRSD and LigW, respectively. SAV2580 converts 1-hydroxy-2-naphthoate to

naphthalene -1-ol with low catalytic constants, k_{cat}, K_m, and k_{cat}/K_m of 0.09 ± 0.01 s^{-1}, 669 ± 109 μM, and 1000 ± 70 M^{-1} s^{-1}; thus, 1-hydroxy-2-naphthoate is probably not the natural substrate of the enzyme. Nevertheless, the presumed substrate has a very similar structure to 1-hydroxy-2-naphthoate. In addition, the high resolution crystal structure (1.75Å) of SAV2580 from *Staphylococcus aureus* in presence of Ca^{2+} has been deposited into the Protein Data Bank (PDB) database under the code number 3NUR. When the active site is closely examined, it is observed that the divalent cation is coordinated by residues found in the ExH motif (Glu-28 and His-30) from β-strand 1, histidine (His-192) from β-strand 5, and aspartate (Asp-316) from β-strand 8 in an octahedral geometry (Figure 4.11). The metal binding residues found in the active site of SAV2580 are virtually identical to the residues found in γ-RSD. Thus, it would be of interest to determine whether SAV2580 would accommodate fused pair of benzene rings rather than a single aromatic ring as seen in the substrates of γRSD.

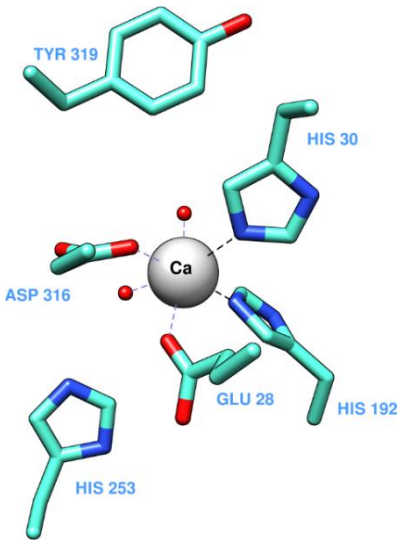


Figure 4.11: Active site of SAV2580 from *Staphylococcus aureus subsp. aureus* Mu50A (PDB id: 3NUR).

CHAPTER V

CONCLUSION

The amidohydrolase superfamily (AHS), identified in 1997 by Holm and Sander, is predominately composed of hydrolytic enzymes, which perform the catalysis at the carbon or phosphorous centers of amino acids, nucleic acids, sugars, and organophosphate esters. All members of AHS fold in distorted $(\beta/\alpha)_8$-barrel with the active site located at the C-terminus of the barrel (*7*). In general, these enzyme could utilize either a mononuclear, binuclear, or trinuclear metal center(s); however, in some case these members function independent of a metal cofactor (*7, 14, 15*). The vast majority of these enzymes however utilize a one or two metal centers for catalysis. Six residues are thought to be important for binding of 1 or 2 metals, which activate the hydrolytic water for catalysis. Besides conservation of active site architecture, little sequence identity is observed among AHS members. The sequence variation is seen in the loop regions of the $(\beta/\alpha)_8$ -TIM barrel, which are thought to undertake conformational changes essential for substrate specificity.

There are 42 different reactions characterized within the AHS, which contains 25,000 members represented by 24 clusters of orthologous groups (COGs) in the NCBI database. These COGs are phylogenetically classified proteins encoded by completely sequenced genomes that are direct evolutionary descendants from a common ancestor (*16*). Members of COG2159 are the only known AHS members that catalyze decarboxylation and hydration reactions. The functions of 7 gene products are known in COG2159, these include γ-resorcylate decarboxylase (γRSD) (*17*), α-amino-β-

carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) (*18*), uracil-5-carboxylate decarboxylase (IDC) (*22*), and 6-methylsalicylic acid decarboxylase (MSD) (*23*), OH-DDVA meta-cleavage compound hydrolase (LigY) (*19*), 4-oxalomesaconate hydratase (LigJ) (*19*), and 5-carboxyvanillate decarboxylase (LigW) (*21*). The last three are gene products of the lignin degrading bacteria *Sphingomonas paucimobilis* SYK-6.

This bacterial strain grows on various lignin related monoaryl and biaryl compounds as a sole carbon and energy source.

Lignin is the most abundant heterogeneous aromatic polymer contained in the structural material of plant cell walls. In nature, the degradation of lignin is cooperative, it commences by various oxidoreductases secreted by white rot fungi. The smaller components are further degraded and mineralized by specialized bacterial enzymes. Microbial lignin degradation systems *in vivo* are extensively studied, although information on the catabolic pathways for breakdown of lignin components is not fully deciphered. In the biphenyl degrader SYK-6, LigY and LigW catalyze two consecutive reactions in the 5, 5'-dehydrodivanillate (DDVA) catabolic pathway and LigJ is found further downstream in the protocatechuate (PCA) 4,5-cleavage pathway. The biphenyl degradation in SYK-6 starts with a model lignin coumpound DDVA, which is converted to 2, 2', 3-trihydroxy-3'-methoxy-5, 5'-dicarboxybiphenyl (OH-DDVA) catalyzed by a demethylase LigX, followed by the formation of a *meta*-cleavage compound by dioxygenase LigZ, which is further hydrolyzed by LigY to 5-carboxyvanillate (5CVA). The last step in the decomposition of DDVA is mediated by a decarboxylase LigW, which converts 5CVA to vanillate (VAN) (*19, 20, 21*). This enzyme is at the interface of two catabolism pathways in the SYK-6 strain, namely the biphenyl degradation and

the PCA 4, 5- cleavage pathway. The latter pathway utilizes VAN up to the formation of pyruvate and oxaloacetate, which in turn enter the TCA cycle (*41*).

Both LigW from *Sphingomonas paucimobilis* SYK-6 and LigW2 from *Novosphingobium aromaticivorans* DSM 12444 catalyze the non-oxidative C-C bond cleavage of 5-CV to VAN (*21*). This transformation is not only important for the degradation of lignin, but it is a key step within the overall degradation pathway for many aromatic compounds. Therefore, it is important to understand how an active site within a superfamily of enzymes that functions primarily to catalyze the hydrolysis of phosphate and carboxylate esters has evolved to catalyze decarboxylation reactions. To better understand the reaction mechanisms of COG2159 decarboxylases various mechanistic probes combined with information on x-ray crystal structures, genomic/operon context, sequence similarity networks (SSN), and substrate library screening have been utilize in this study.

The high resolution crystal structures of LigW and LigW2 in the presence of Mn^{2+} and various substrate analogues have defined the substrate binding mode into the active site and the substrate conversion into the transition state. The 1.07 Å high-resolution structure of LigW2 in complex with Mn^{2+} and a very close unreactive substrate analogue and a tight-binding inhibitor, 5-nitrovanillate (5-NV), have shown the analogue coordinated in a bidentate fashion to the metal center. In this structure, the nitro and phenolate groups of 5-NV are coordinated directly to the single Mn^{2+} cation and together with a glutamate from β-strand 1, a histidine from β-strand 5, and an aspartate from β-strand 8 retain an octahedral geometry. The latter two residues are highly conserved throughout all the members of COG2159 that catalyze a

decarboxylation reaction. However, in LigW the glutamate from β-strand 1 belongs to an ExA motif, which is not conserved in all the decarboxylases and deviates from the typical HxH motif found in many AHS member. In the LigW2 structure complexed with 5-NV (PDB id: 4QRN) the bidentate ligand displays a distorted planar geometry, which is also observed in the structure of LigW with the same inhibitor (PDB id: 4QRN). However, the near angstrom structure of LigW2 reveals that the hydroxyl group and the nitro group of 5-NV are ~13^0 and ~23^0 off the main benzene ring plane, similarly observation is made in the LigW structure, where the nitro group is also out of plane by ~22° and the hydroxyl by ~11°. In addition, the crystal structures of LigW in presence of vanillate (VAN) and 3-methoxy-5-carboxybenzoate (MCB), respectively show similar distortions of the substrate analogues (PDB ids: 4L6D and 4NI8). The nitro inhibitor was further crystalized as a free compound indicating that the molecule is indeed planer in solution. Comparison of this structure with the enzyme bound inhibitor supports the notion that the distortion of 5NV occurs in the process of binding inside the rigid enzyme active site.

The overall topology of LigW and LigW2 is similar to other COG2159 decarboxylases with exception of an insertion domain found between β-strands 1 and 2. These insertion domains are unique for each structurally characterized decarboxylase and are crucial for substrate binding and orientation in the active sites. In ACMSD (PDB id: 4IH3) and IDC (PDB id: 4HK6), the insertion domains are composed of a β-sheet and β-hairpin, respectively. However, both domains contain arginine residues, which interact with the substrates. In addition, γRSD (PDB id: 2DVU) contains an arginine and a phenylalanine within insertion domain composed of a single α-helix.

Whereas, in LigW and LigW2 the insertion domain is composed of three α-helices. Interestingly, a single arginine from the insertion domain in LigW maintains strong polar contacts with the carboxylate group at C-1 of 5NV. Whereas in LigW2, an additional tyrosine also from the insertion domain helps stabilize the C-1 carboxylate of the inhibitor. Furthermore, an invariant arginine residue from the adjacent monomer in LigW and LigW2 participates in ionic interactions with the nitro group. This residue is highly conserved among all the decarboxylases in COG2159, suggesting that the minimal catalytic form of these enzymes is a homodimer. The enzymes catalyzing the decarboxylation of a molecule with an aromatic ring, typically contain hydrophobic residue that stabilizes the ring through π-π stacking interactions. In LigW, LigW2, γRSD, and IDC, these residues are Tyr-299, Tyr-317, Phe-290, and Phe-326, respectively. Therefore, the information of these traits could be utilized, when predicting functions of hypothetically decarboxylases in COG2159.

Mutations of the active site residues in LigW important for metal or substrate binding (Glu-7, Arg-45, His-173, Tyr-197, His-226, Glu-229, Arg-237, Asp-296, Tyr-299) diminished the k_{cat} and k_{cat}/K_m by at least three orders of magnitude. The pH profile of LigW and LigW2 is bell shaped indicated the required deprotonation and protonation of two residues, the highly conserved aspartate from β-strand 8 and histidine from β-strand 6, respectively. The latter is part of a HxxE motif and participates in a hydrogen bond triad with the aspartate from β-strand 8. Presumably these residues could facilitate the shuttling of protons to the active site or modulate the pK_a value of the aspartate from β-strand 8. The x-ray structures of LigW and LigW2, support the notion that the hydrogen at C-5 of the product vanillate is donated either from the metal bound aspartate

or the phenolic group at C-4 of 5-CV. The closest other entity that could function as the proton donor is the phenolic substituent from C4. Since, Asp-296 is conserved in all characterized AHS members and could function as a general acid, this residue is likely the proton donor during catalysis. Mutation of this residue to an asparagine in LigW results in the reduction in catalytic activity by more than three orders of magnitude and thus this residue is critical for the catalytic activity of LigW. The out-of-plane bending observed in the crystal structures with 5-NV would facilitate the protonation of C-5 from the *si*-face from either the hydroxyl group at C4 or the carboxylate group of the aspartate from β-strand 8. If protonation at C-5 occurs after decarboxylation, it is anticipated that this step would be fast relative to the actual decarboxylation step and thus the product isotope effect (PIE), determined using a 50:50 mixture of H_2O:D_2O, should be near unity. However, with LigW the PIE was determined to be 4.6 and thus LigW catalyzes the decarboxylation with a proton transfer to the C-5 of the benzene ring prior to removal of the CO_2 group. Furthermore, it has also been demonstrated that LigW will catalyze the exchange of the proton at C-5 of the product vanillate with a rate constant of 7.9×10^{-3} s^{-1} at pD 9.0. However, no exchange with solvent over an incubation time of 7 days in D_2O using 3-methoxybenzoate was observed. Therefore, the exchange reaction requires the participation of the phenolic group at C-4 via the delocalization of the lone pair of electrons from the phenolate to C-5. In addition, it was determined that a water molecule is not utilized by LigW for the conversion of 5-CV to vanillate. In the nitro inhibitor bound structure it is clear that the remaining metal bound water is ill positioned and not in a proximity to attack the aromatic ring or the C5-carboxylate group. Furthermore, the initial product of the reaction catalyzed by LigW is in fact CO_2 rather

than bicarbonate as supported by the findings from the membrane inlet mass spectrometry (MIMS) experiment.

The chemical mechanism of decarboxylation of 5-CV to VAN by LigW has been proposed. Upon binding of 5-CV into the active site, two water molecules are displaced and the phenolic group at C4 and the carboxylate oxygen at C-5 coordinate to the mononuclear metal center in a bidentate fashion retaining the Mn^{2+} preferred octahedral geometry. The binding of the substrate into the active site is accompanied by a distorted geometry and a reduction of the pK_a value of the phenolic group at C4. This facilitates a proton from the phenolic group is either transferred to the aspartate from β-strand 8 or directly to the C-5 of the benzene ring after delocalization of the lone pair. The proton transfer is likely to be rate limiting as supported by the significant solvent and product isotope effects. Decarboxylation then occurs with subsequent release of the product from the active site.

γ-Resorcylate decarboxylase (γRSD) is the closest structural homologue of LigW and similarly utilizes a hydroxybenzoate as a substrate. The enzyme is found in the degradation pathways of γ-resorcylate and it catalyzes the first step by decarboxylating the compound to resorcinol in a nonoxidative fashion. The γ-RSD from *Polaromonas* sp. JS666 also facilitates the decarboxylation reactions of 2, 3-dihydroxybenzoate, 2, 4, 6-trihydroxybenzoate, and 2, 6-dihydroxy-4-methylbenzoate. The reaction of γ-resorcylate by γRSD was inhibited in the presence of a strong inhibitor 2-nitroresorcinol (2-NR). The crystal structure of γRSD in complex with Mn^{2+} and 2-NR have revealed that the binding mode and distorted geometry of the substrate analogue closely resembles that of 5-NV in LigW. Furthermore, a hydrogen bond triad involving HxxE

motif with the metal bound aspartate is also seen in the γRSD crystal structure and the invariant arginine from the adjacent subunit interacts with the nitro group as seen in LigW, 5-NV structure. The metal content of γRSD has not been fully elucidated and the amount of metal necessary for optimal turnover and efficiency has not been determined. In search of new functions in COG2159, a hypothetical decarboxylase, SAV2580 from *Staphylococcus aureus subsp. aureus Mu50A,* was screened with a library of potential substrates. It was determined that the enzyme had a slight activity towards the conversion of 1-hydroxy-2-naphthoate to naphthalene-1-ol. The natural substrate for this enzyme remains to be discovered.

By elucidating the reaction mechanism of LigW, we have taken a step closer to understanding how a group of enzymes, which are predominately composed by hydrolases have evolved into decarboxylases.

REFERENCES

1. The UniProt consortium. (2015) UniProt: a hub for protein information, *Nucl. Acids Res. 43*, D204-D212.
2. Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies, *PLoS Comput. Biol. 5*, e10000605.
3. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D., and Maltsev, N. (1999) The use of gene clusters to infer functional coupling, *Proc. Natl. Acad. Sci. 96*, 2896-2901.
4. Price, M., Haung, K. H., Arkin, A. P., and Alm, E. J. (2005) Operon formation is driven by co-regulation and not by horizontal gene transfer, *Genome Res. 15*, 809-819.
5. Piper, U., Chiang, R., Seffernick, J. J., Brown, S. D., Glasner, M. E., Kelly, L., Eswar, N., Sauder, J. M., Bonanno, J. B., Swamninathan, S., Burley, S. K., Zheng, X., Chance, M. R., Almo, S. C., Gerlt, J. A., Raushel, F. M., Jacobson, M. P., Babbitt, P. C., and Sali, A. (2009) Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamily, *J. Struct. Funct. Genomics 10*, 107-125.
6. Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., Dunaway-Mariano, D., Imker, H. J., Jacobson, M. P., Minor, W., Poulter, C. D., Raushel, F. M., Sali, A., Shoichet, B. K., and Sweedler, J. V. (2011) The enzyme function initiative, *Biochemistry 50*, 9950-9962.

7. Seibert, C. M. and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry 44*, 6383-6391.

8. Raushel, F. M. (2009) Functional annotation of orphan enzymes within the amidohydrolase superfamily, *Beilstein-Institut.*, 9-19.

9. Hall, R. S., Xiang, D. F., Xu, C., and Raushel, F. M. (2007) *N*-Acetyl-D-glucosamine-6-phosphate deacetylase: substrate activation via a single divalent metal ion, *Biochemistry 46*, 7942-7952.

10. Bigley, A. N and Raushel, F. M. (2013) Catalytic mechanisms for phosphotriesterases. *Biochim Biophys Acta.* 1834, 443-453.

11. Thoden, J. B., Phillips, Jr. G. N., Neal, T. N., Raushel, F. M., and Holden, H. M. (2001) Molecular structure of dihydroorotase: a paradigm for catalysis through the use of a binuclear metal center, *Biochemistry 40*, 6989-6997.

12. Holm, L. and Sander, C. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease, *Proteins 28*, 72-82

13. Hall, R. S., Brown, S., Fedorov, A. A., Fedorov, E. V., Xu, C., Babbitt, P. C., Almo, S. C., and Raushel, F. M. (2007) Structural diversity within the mononuclear and binuclear active sites of *N*-acetyl-D-glucosamine-6-phosphate deacetylase, *Biochemistry 46*, 7953-7962.

14. Hobbs, M. E., Malashkevich, V., Williams, H. J., Xu, C., Sauder, M. J., Burley, S. K., Almo, S. C., and Raushel, F. M. (2012) Structure and catalytic mechanism of LigI: insight into the amidohydrolase enzymes of cog3618 and lignin degradation, *Biochemistry 51*, 3497-3507.

15. Ghodge, S. V., Fedorov, A. A., Fedorov, E. V., Hillerich, B., Seidel, R., Almo, S. C., and Raushel, F. M. (2013) Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the polymerase and histidinol phosphatase family of proteins, *Biochemistry 52*, 1101-1112.

16. Tatusov, R. L., Galperin, M. Y., Natale, D. A. and Koonin, E. V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution, *Nucleic Acid Research. 28*, 33-36.

17. Yoshida, T., Hayakawa, Y., Matsui, T., and Nagasawa, T. (2004) Purification and characterization of 2,6-dihydroxybenzoate decarboxylase reversibly catalyzing nonoxidative decarboxylation, *Arch. Microbiol. 181*, 391-397.

18. Li, T., Iwaki, H., Fu, R., Hasegawa, Y., Zhang, H., and Liu, A. (2006) α-amino-β-carboxymuconic-ε-semialdehyde decarboxylase (ACMSD) is a new member of the amidohydrolase superfamily, *Biochemistry 45*, 6628-6634.

19. Peng, X., Masai, E., Katayama, Y. and Fukuda, M. (1999) Characterization of *meta*-cleavage compound hydrolase gene involved in degradation of the lignin-related biphenyl structure by *Sphingomonas paucimobilis* SYK-6, *Appl. Environ. Microbiol*. *65*, 2789-2793.

20. Hara, H., Masai, E., Katayama, Y. and Fukuda, M. (2000) The 4-Oxalomesaconate hydratase gene, involved in the protocatechuate 4,5-cleavage pathway, is essential to vanillate and syringate degradation in *Sphingomonas paucimobilis* SYK-6, *J. Bacteriol.* 182, 6950-6957.

21. Peng, X., Masai, E; Kitayama, H., Harada, K., Katayama, Y. and Fukuda, M. (2002) Characterization of the 5-carboxyvanillate decarboxylase gene and its role

in lignin-related biphenyl catabolism in *Sphingomonas paucimobilis* SYK-6, *Appl. Environ. Microbiol.* 68, 4407-4415.

22. Xu, S., Li, W., Zhu, J., Wang, R., Li, Z., Xu, G., and Ding, J. (2013) Crystal structure of isoorotate decarboxylase reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylases, *Cell Research. 23,* 1296-1309.

23. Snini, S. P., Tadrist, S., Lafitte, J., Jamin, J., Jamin, E. L., Oswald, I. P., and Puel, O. (2014) The gene patG involved in the biosynthesis pathway of patulin, a food-borne mycotoxin, encodes a 6-methylsalicylic acid decarboxylase, *Int. J. Food Microbiol. 171,* 77-83.

24. Gato, M., Hayashi, H., Miyahara, I., Hirotsu, K., Yoshida, M. and Oikawa, T. (2006) Crystal structures of nonoxidative zinc-dependent 2,6-dihydroxybenzoate (γ-resorcylate) decarboxylase form *Rhizobium* sp. Strain MTP-100005, *J.Biol. Chem. 281,* 34365-34373.

25. Garavaglia, S., Perozzi, S., Galeazzi, L., Raffaelli, N., and Rizzi, M. (2009) The crystal structure of human α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase in complex with 1,3-dihydroxyacetonephosphate suggests a regulatory link between NAD synthesis and glycolysis, *FEBS J. 276,* 6615-6623.

26. Martynowski, D., Eyobo, Y., Li, T., Yang, K., Liu, A., and Zhang, H. (2006) Crystal structure α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase: insight into the active site and catalytic mechanism of a novel decarboxylation reaction, *Biochemistry 45,* 10412-10421.

27. Huo, L., Liu, F., Iwaki, H., Li, T., Hasegawa, Y., and Liu, A. (2015) Human α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD): a structural and mechanistic unveiling. *Proteins*. *83,* 178–187.

28. Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Capilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A. R.,Vailaya, A., Wang, P., Adler, A., Conklin, B. R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G. J., Ideker, T. and Bader, G. D. (2007) Integration of biological networks and gene expression data using cytoscape, *Nature Protocols 2,* 2366 – 2382.

29. Atkinson, H. J., Morris, J. H., Ferrin, T. E. and Babbitt, P. C. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies, *PLoS Comput. Biol. 4*, e4345.

30. Tatusov, R. L., Galperin, M. Y., Natale, D. A. and Koonin, E. V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution, *Nucleic Acid Research 28,* 33-36.

31. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families, *Science 278,* 631-637.

32. Korbel, J. O., Jensen, L. J., Mering, C. V., and Bork, P. (2004) Analysis of genomic context: prediction of functional associations form conserved bidirectionally transcribed gene pairs, *Nature Biotechnol. 22,* 911-917.

33. Masai, E., Y. Katayama, S. Nishikawa, and M. Fukuda (1999) Characterization of *Sphingomonas paucimobilis* SYK-6 genes involved in degradation of lignin-related compounds, *J. Ind. Microbiol. Biotechnol. 23*, 364-373.

34. Masai, E., Harada, K., Peng, X., Kitayama, H., Katayama, Y., and Fukuda, M. Cloning and characterization of the ferulic acid catabolic genes of *Sphingomonas paucimobilis* SYK-6, *Appl Environ Microbiol. 68*, 4416-4424.

35. Xu, C., Arancon, R. A. D., Labidi, J., and Luque, R. (2014) Lignin depolymerisation strategies: towards valuable chemicals and fuels, *Chem. Soc. Rev. 43*, 7485-7500.

36. Tobimatsu, Y., Chen, F., Nakashima, J., Escamilla-Treviño, L. L., Jackson, L., Dixon, R. A., and Ralph, J. (2013) Coexistence but independent biosynthesis of catechyl and guaiacyl/syringyl lignin polymers in seed coats, *Plant Cell. 25*, 2587-2600.

37. Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Rahmanpour, R. (2011) Pathways for degradation of lignin in bacteria and fungi, *Nat. Prod. Rep. 28*, 1883-1896.

38. Sainsbury, P. D., Hardiman, E. M., Ahmad, M., Otani, H., Seghezzi, N., Eltis, L. D., and Bugg, T. D. H. (2013) Breaking down lignin to high-value chemicals: the conversion of lignocellulose to vanillin in a gene deletion mutant of *Rhodococcus jostii* RHA1, *ACS Chem. Biol. 8*, 2151-2156.

39. Tuor, U., Winterhalter, K., and Fiechter, A. (1995) Enzymes of white-rot fungi involved in lignin degradation and ecological determinants for wood decay, *J. Biotechnol. 41*, 1-17.

40. Abdel-Hamid, A. M., Solbiati, J. O., and Cann, I.K. (2013) Chapter One – Insights into lignin degradation and its potential industrial applications, *Adv. Appl. Microbiol. 82*, 1-28.

41. Masai, E., Katayama, Y., and Fukuda, M. (2007) Genetic and biochemical investigation of bacterial and catabolic pathways for lignin-derived aromatic compounds, *Biosci. Biotechnol. Biochem. 71*, 1-15.

42. Sonoki, T., Masai, E., Sato, K., Kajita, S., and Katayama Y. (2009) Methoxyl groups of lignin are essential carbon donors in C1 metabolism of *Sphingobium* sp. SYK-6, *J. Basic Microbiol 49*, S98-S102.

43. Kasai, D., Masai, E., Miyauchi, K., Katayama, Y., and Fukuda M. (2004) Characterization of the 3-O-methylgallate dioxygenase gene and evidence of multiple 3-O-methylgallate catabolic pathways in *Sphingomonas paucimobilis* SYK-6, *J Bacteriol. 186*, 4951-4959.

44. Li, T., Huo, L., Pulley, C., and Liu, A. (2012) Decarboxylation mechanisms in biological systems, *Bioorg. Chem. 43*, 2-14.

45. Peng, X., Masai, E., Kasai, D., Miyauchi, K., Katayama, Y. and Fukuda, M. (2005) A second 5-carboxyvanillate decarboxylase gene, ligW2, is important for lignin-related biphenyl catabolism in *Sphingomonas paucimobilis* SYK-6. *Appl. Environ. Microbiol. 71*, 5014-5021.

46. Takeuchi, M., Hamana, K., and Hiraishi, A. (2001) Proposal of the genus *Sphingomonas sensu stricto* and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses, *Int. J. Syst. Evol. Microbiol. 51*, 1405-1417.

47. Lyu, Y., Zheng, W., Zheng, T, and Tian. Yun. (2014) Biodegradation of Polycyclic Aromatic Hydrocarbons by *Novosphingobium pentaromativorans* US6-1, *PLoS ONE 9*, e101438.

48. Kertesz, M. A. and Kawasaki, A. (2010) Hydrocarbon-degrading Sphingomonads: Sphingomonas, Sphingobium, Novosphingobium, and Sphingopyxis, *Handbook of Hydrocarbon and Lipid Microbiology*, 1693-1705.

49. Danzer, C. and Mattner, J. (2013) Impact of microbes on autoimmune diseases, *Arch. Immunol. Ther. Exp. (Warsz). 61*, 175-186.

50. Habe, H. and Omori, T. (2003) Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria, *Biosci. Biotechnol. Biochem*. *67*, 225–243.

51. Kawahara, K., Moll, H., Knirel, Y. A, Seydel, U, and Zahringer, U. (200) Structural analysis of two glycosphingolipids from the lipopolysaccharide-lacking bacterium *Sphingomonas capsulata*. *Eur. J. Biochem. 267*, 1837–1846.

52. Fredrickson, J.K., Brockman, F. J., Workman, D. J., Li, S. W., and Stevens, T. O. (1991) Isolation and characterization of a subsurface bacterium capable of growth on toluene, naphthalene, and other aromatic compounds, *Appl. Environ. Microbiol. 57*, 796-803.

53. Peng. X., Egashira, T., Hanashiro, K., Masai, E., Nishikawa, S., Katayama, Y., Kimbara, K., and Fukuda, M. (1998) Cloning of a *Sphingomonas paucimobilis* SYK-6 gene encoding a novel oxygenase that cleaves lignin-related biphenyl and characterization of the enzyme, *Appl. Environ. Microbiol. 64*, 2520-2527.

54. Kaplan, M.M. (2004) *Novosphingobium aromaticivorans*: a potential initiator of primary biliary cirrhosis, *Am. J. Gastroenterol. 99*, 2147-9.

55. Gerlt, J. A. and Raushel, F. M. Evolution of function in (beta/alpha)$_8$-barrel enzymes. *Curr. Opin. Chem. Biol. 7*, 252-264.

56. Kiss, L. E., Ferreira, H.S., Torrao, L., Bonifácio, M. J., Palma, P. N., Soares-da-Silva, P., and Learmonth, D. A. Discovery of a long-acting, peripherally selective inhibitor of catechol-O-methyltransferase, *J. Med. Chem. 53*, 3396-3411.

57. Hatfield, R. and Vermerris, W. (2001) Lignin formation in plants. The dilemma of linkage specificity, *Plant Physiology 126*, 1351-1357.

58. Ruttimann, C., Vicuna, R., Mozuch, M. D., and Kirk, T. K., (1991) Limited bacterial mineralization of fungal degradation intermediates from synthetic lignin, *Appl. Environ. Microbiol. 57*, 3652-3655.

59. Marti-Arbona, R., Xu, C., Steele, S., Weeks, A., Kuty, G. F. Seibert, C. M., and Raushel, F. M. (2006) Annotating enzymes of unknown function: *N*-formimino-L-glutamate deiminase is a member of the amidohydrolase superfamily, *Biochemistry 45*, 1997-2005.

60. Yoshida, M., Fukuhara, N., and Oikawa, T. (2004) Thermophilic, reversible gamma-resorcylate decarboxylase from *Rhizobium* sp. strain MTP-10005: purification, molecular characterization, and expression, *J. Bacteriol. 186*, 6855-6863.

61. Muraki, T., Taki, M., Hasegawa, Y., Iwaki, H., and Lau, P. C. K. (2003) Prokaryotic homologs of the eukaryotic 3-hydroxyanthranilate 3,4-dioxygenase and 2-amino-3-carboxymuconate-6-semialdehyde decarboxylase in the 2-

nitrobenzoate degradation pathway of *Pseudomonas fluorescens* strain KU-7, *Appl. Environ. Microbiol.69*, 1564-1572.

62. Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., Dunaway-Mariano, D., Imker, H. J., Jacobson, M. P., Minor, W., Poulter, C. D., Raushel, F. M., Sali, A., Shoichet, B. K., and Sweedler, J. V. (2011) The enzyme function initiative, *Biochemistry 50*, 9950-9962.

63. Xu, C., Hall, R., Cummings, J. and Raushel, F. M. (2006) Tight binding inhibitors of *N*-acyl amino sugar and *N*-acyl amino acid ddeacetylase, *J. Am. Cham. Soc. 128*, 4244-4245.

64. Krężel, A., and Bal, W. (2004) A formula for correlating pK_a values determined in D_2O and H_2O, *J. Inorg. Biochem. 98*, 161-6.

65. Moral, M. E.G., Tu, C., Richards, N. G. J., and Silverman, D. N. (2011) Membrane inlet for mass spectrometric measurement of catalysis by enzymatic decarboxylases, *Analy. Biochem. 418*, 73–77.

66. Mundle, S. O. C., Lacrampe-Couloume, G., Lollar, B. S., Kluger, R. (2010) Hydrolytic decarboxylation of carboxylic acids and the formation of protonated carbonic acid. *J. Am. Chem. Soc. 132*, 2430-2436.

67. Toth, K., Amyes, T. L., Wood, B. M., Chan, K., Gerlt, J. A., and Richard, J. P. (2010) Product deuterium isotope effects for orotidine 5'-monophosphate decarboxylase: effect of changing substrate and enzyme structure on the partitioning of the vinyl carbanion reaction intermediate, *J. Am. Chem. Soc. 1*32, 7018-7024.

68. Amyes, T. L., Wood, B. M., Chan, K., Gerlt, J. A., and Richard, J. P. (2008) Formation and stability of a vinyl carbanion at the active site of orotidine 5'-monophosphate decarboxylase: pKa of the C-6 proton of enzyme-bound UMP, *J. Am. Chem. Soc. 130,* 1574-1575.

69. Liu, A. and Zhang, H. (2006) Transition metal-catalyzed nonoxidative decarboxylation reactions, *Biochemistry 45,* 10407-10411.

70. Huo, L., Fielding, A. J., Chen, Y., Tingfeng, L., Iwaki, H., Hosler, J. P., Chen, L., Hasegawa, Y., Que, Jr., L., and Liu, A. (2012) Evidence for a dual role of an active site histidine in α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase, *Biochemistry 51,* 5811-5821.

71. Huo, L., Davis, I., Chen, L. and Liu, A. (2013) The power of two: arginine 51 and arginine 239* from a neighboring subunit are essential for catalysis in α-amino-β-carboxymuconate-epsilon-semialdehyde decarboxylase, *J. Biol. Chem. 288,* 30862-30871.

72. Iwasaki, Y., Kino, K., Nishide, H., and Kirimura, K. (2007) Regioselective and enzymatic production of γ-resorcylic acid from resorcinol using recombinant *Escherichia coli* cells expressing a novel decarboxylase gene, *Biotechnol Lett.29,* 819-822.

73. Kluge, C., Tschech, A., and Fuchs, G. (1999) Anaerobic metabolism of resorcylic acids (m-dihydroxybenzoic acids) and resorcinol (1, 3-benzenediol) in fermenting and in a denitrifying bacterium, *Arch. Microbiol. 155,* 68-74.

74. Yoshida, M., Oikawa, T., Obata, H., Abe, K., Mihara, H., and Esaki, N. (2007) Biochemical and genetic analysis of the γ-resorcylate (2,6-dihydroxybenzoate)

catabolic pathway in *Rhizobium* sp. strain MTP-10005: identification and functional analysis of its gene cluster, *J. Bacteriol. 189*, 1573-1581.

75. Ishii, Y., Narimatsu, Y., Iwasaki, Y., Arai, N., Kino, K., and Kirimura, K. (2004) Reversible and nonoxidative γ-resorcylic acid decarboxylase: characterization and gene cloning of a novel enzyme catalyzing carboxylation of resorcinol, 1,3-dihydroxybenzene, from *Rhizobium radiobacter, Biochem. Biophys. Res. Commun. 324*, 611-620.

76. Li, T., Huo, L, Pulley, C., and Liu, A. (2012) Decarboxylation mechanisms in biological systems, *Bioorg. Chem. 43*, 2-14.

77. Iiams, V., Desai, B. J., Fedorov, A. A., Fedorov, A. A., Almo, S. C., and Gerlt, J. A. (2011) Mechanism of the orotidine 5'-monophosphate decarboxylase-catalyzed reaction: importance of residues in the orotate binding site, *Biochemistry 50*, 8497-8507.

78. Degrassi, G., Polverino De Laureto, P. and Bruschi, C. V. (1995) Purification and characterization of ferulate and p-coumarate decarboxylase from *Bacillus pumilus*, *Appl. Environ. Microbiol. 61*, 326-332.

79. Huder, J. B. and Dimroth, P. (1995) Expression of the sodium ion pump methylmalonyl-coenzyme a decarboxylase from *Veillonella parvula* and of mutated enzyme specimens in *Escherichia coli*, *J. Bacteriol. 177*, 3623-3630.

80. Wuensch, C., Gross, J., Steinkellner, G., Lyskowski, A., Gruber, K, Glueck, S. M., and Faber, K. (2014) Regioselective ortho-carboxylation of phenols catalyzed by benzoic acid decarboxylases: a biocatalytic equivalent to the Kolbe–Schmitt reaction, *RSC Adv. 4*, 9673-9679.

81. Wensch, C., Schmidt, N., Gross, J., Grischek, B., Glueck, S. M., and Faber. (2013) Pushing the equilibrium of regio-complementary carboxylation of phenols and hydroxystyrene derivatives, *J. Biotechnol. 168*, 264-270.

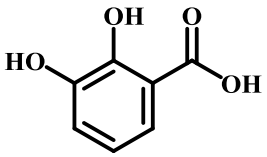
82. Reid, J., Watson, R. D., Cochran, J. B., Sproull, D. H., Clayton, B. E., and Prunty, F. T. (1951) Sodium γ-resorcylate in rheumatic fever, *Br. Med. J. 4727*, 321-326.

83. Santha, R., Savithri, H. S., Rao, A., and Vaidyanathan, C. S. 2,3-(1995) Dihydroxybenzoic acid decarboxylase from *Aspergillus niger*: a novel decarboxylase, *Eur. J. Biochem. 230*, 104-110.

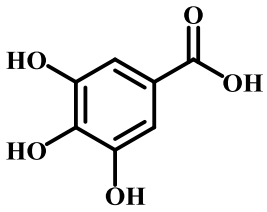
84. Kirimura, K., Gunji, H., Wakayama, R., Hattori, T., and Ishii, Y. (2010) Enzymatic Kolbe-Schmitt reaction to form salicylic acid from phenol: enzymatic characterization and gene identification of a novel enzyme, *Trichosporon moniliiforme* salicylic acid decarboxylase *Biochem. Biophys. Res. Commun. 394*, 279-284.

APPENDIX

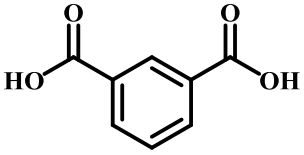
Below are the chemical structures of the compounds tested as potential substrates of γRSD from *Polaromonas* sp. JS666 and SAV2580 from *Staphylococcus aureus subsp. aureus* Mu50.



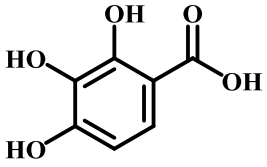
2,3-dihydroxybenzoic acid



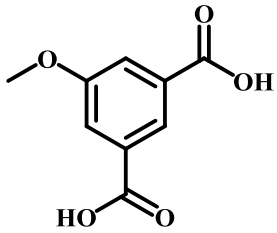
3,4,5-trihydroxybenzoic acid



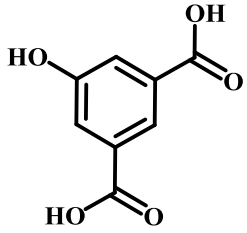
isophthalic acid



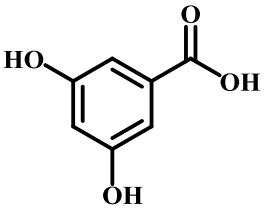
2,3,4-trihydroxybenzoic acid



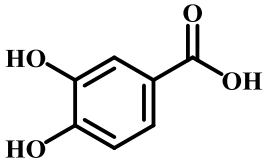
5-methoxyisophthalic acid



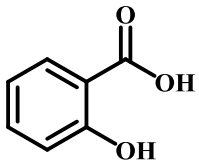
5-hydroxyisophthalic acid



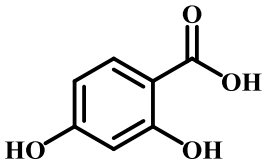
3,5-dihydroxybenzoic acid



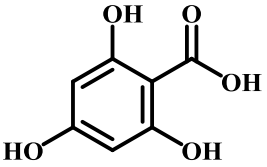
3,4-dihydroxybenzoic acid



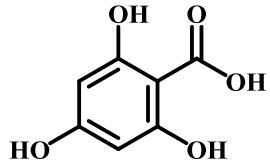
2-hydroxybenzoic acid



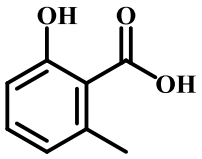
2,4-dihydroxybenzoic acid



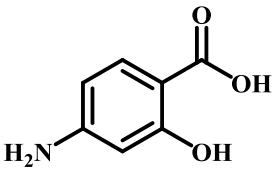
2,4,6-trihydroxybenzoic acid



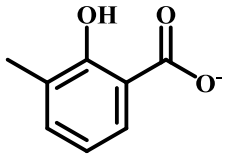
2,4,6-trihydroxybenzoic acid



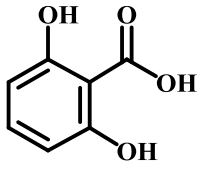
6-methylsalicylate



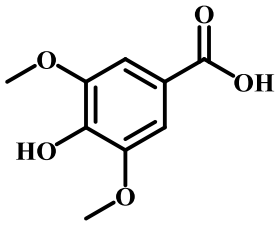
4-aminosalicylate



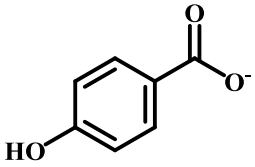
3-methylsalicylate



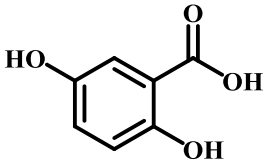
2,6-dihydroxybenzoic acid



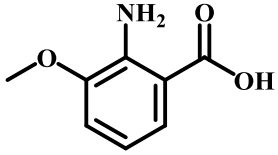
4-hydroxy-3,5-dimethoxybenzoic acid



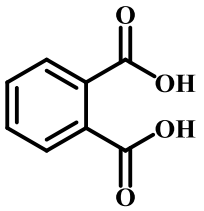
4-hydroxybenzoate



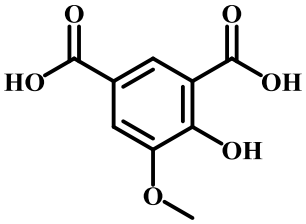
2,5-dihydroxybenzoic acid



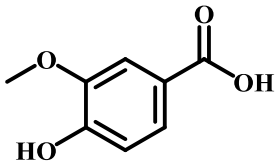
2-amino-3-methoxybenzoic acid



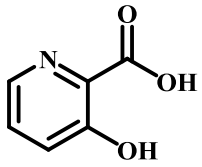
phthalic acid



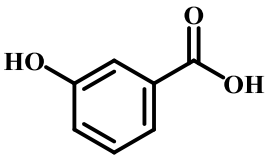
4-hydroxy-5-methoxyisophthalic acid



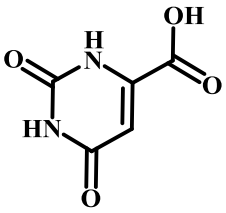
4-hydroxy-3-methoxybenzoic acid



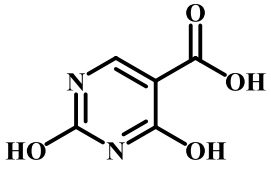
3-hydroxypicolinic acid



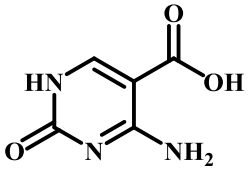
3-hydroxybenzoic acid



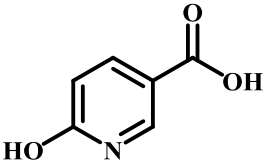
orotic acid



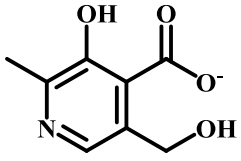
2,4-Dihydroxypyrimidine-5-carboxylic acid



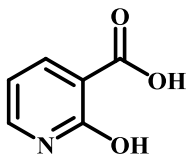
5-carboxycytosine



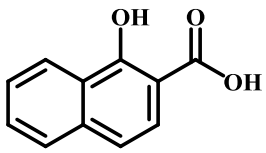
6-hydroxynicotinic acid



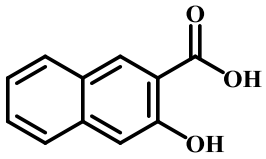
4-pyridoxate



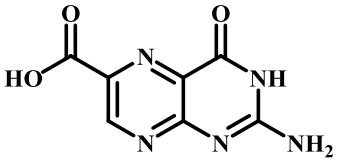
2-hydroxynicotinic acid



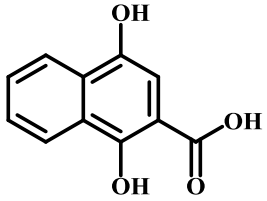
1-hydroxy-2-naphthoic acid



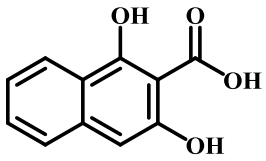
3-hydroxy-2-naphthoic acid



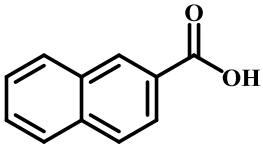
pterine-6-carboxylic acid



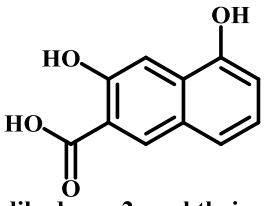
1,4-dihydroxy-2-naphthoic acid



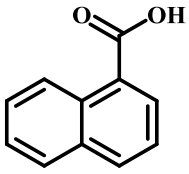
1,3-dihydroxy-2-naphthoic acid



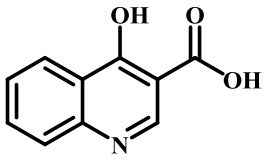
2-naphthoic acid



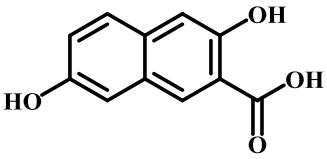
3,5-dihydroxy-2-naphthoic acid



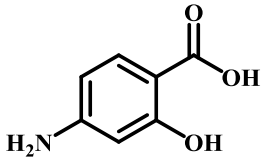
1-naphthoic acid



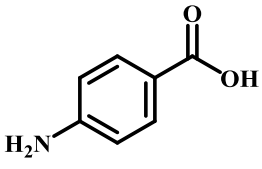
4-hydroxyquinoline
-3-carboxylic acid



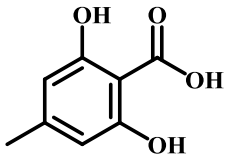
3,7-dihydroxy-2-naphthoic acid



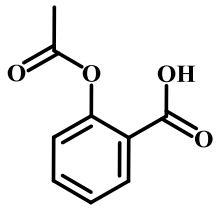
4-amino-2-hydroxybenzoic acid



4-aminobenzoic acid



2,6-dihydroxy-4-methylbenzoic acid



2-acetoxybenzoic acid